

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization  
International Bureau



(43) International Publication Date  
22 November 2001 (22.11.2001)

PCT

(10) International Publication Number  
WO 01/88114 A2

(51) International Patent Classification<sup>2</sup>: C12N 15/00

(21) International Application Number: PCT/EP01/05794

(22) International Filing Date: 18 May 2001 (18.05.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:  
0012231.7 19 May 2000 (19.05.2000) GB

(71) Applicant (for all designated States except US): DEV-  
GEN NV [BE/BE]; Technologiepark 9, B-9052 Zwij-  
naarde (BE).

(72) Inventors; and

(75) Inventors/Applicants (for US only): PLAETINCK,  
Geert [BE/BE]; Pontstraat 16, B-9820 Merelbeke (BE).  
RENARD, Jean-Pierre [BE/BE]; Peter Benoitlaan  
141, B-9050 Gentbrugge (BE). MORTIER, Katherine  
[BE/BE]; Lange Boomgaardstraat 25, B- Gent (BE).  
BOGAERT, Thierry [BE/BE]; Wolvendreef 26g, B-8500  
Kortrijk (BE).

(74) Agent: BOULT WADE TENNANT; Verulam Gardens,  
70 Gray's Inn Road, London WC1X 8BT (GB).

(81) Designated States (national): AE, AG, AL, AM, AT, AU,  
AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU,  
CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM,  
HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK,

LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX,  
MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL,  
TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States (regional): ARIPO patent (GH, GM,  
KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian  
patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European  
patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE,  
IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF,  
CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Declarations under Rule 4.17:

- as to applicant's entitlement to apply for and be granted a  
patent (Rule 4.17(ii)) for the following designations AE, AG,  
AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN,  
CO, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD,  
GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR,  
KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN,  
MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI,  
SK, SL, TJ, TM, TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZW,  
ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ,  
UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU,  
TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES,  
FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI  
patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR,  
NE, SN, TD, TG)
- of inventorship (Rule 4.17(iv)) for US only

Published:

- without international search report and to be republished  
upon receipt of that report

For two-letter codes and other abbreviations, refer to the "Guid-  
ance Notes on Codes and Abbreviations" appearing at the begin-  
ning of each regular issue of the PCT Gazette.

(54) Title: GENE EXPRESSION SYSTEM

(57) Abstract: There are provided DNA constructs, including replicable cloning vectors and expression vectors, comprising a bac-  
teriophage promoter operably linked to an outon sequence. The expression vectors provided by the invention are useful in the  
expression of recombinant polypeptides in host cells or organisms and are particularly useful in expression of recombinant polypep-  
tides in nematode worms such as *C. elegans*.

WO 01/88114 A2

NUC 38884

- 1 -

## GENE EXPRESSION SYSTEM

## Field of the invention

5           The invention relates to the expression of DNA, genes, cDNAs, proteins, peptides and parts thereof in the nematode worm *C. elegans*. In particular, the invention relates to methods of improving the translation of RNAs transcribed in *C. elegans* using a  
10 bacteriophage polymerase by introduction of a trans-splice recognition site recognised by an SL1 trans-splice recognition sequence into the DNA template transcribed by the bacteriophage polymerase.

## 15   Background to the invention

Eukaryotic versus prokaryotic expression.

          Bacteriophage RNA polymerases, such as T7, T3, and SP6, and their corresponding promoters have been  
20 used extensively to drive the expression of heterologous genes in a variety of organisms. In co-pending International patent application No. WO 00/01846, Plaetinck et al. describe the use of the T7 system to express DNA, genes, cDNA, proteins and  
25 peptides of parts thereof and for the expression of double-stranded RNA (dsRNA) in the nematode model system *C. elegans*.

          The bacteriophage expression systems are well known in the art for use in prokaryotic host cells,  
30 such as *E. coli*, and have the advantage that they provide simple and strong expression systems dependent only on one RNA polymerase and one well defined promoter. The application of such efficient expression systems in eukaryotic organisms is,  
35 however, not evident, mainly because messenger RNAs from eukaryotes and prokaryotes have a different

- 2 -

structure, which has implications for translation efficiency and RNA stability.

1 Messenger RNAs of higher eukaryotes share a functionally essential 5' CAP structure. This  
5 structure is generated during a capping reaction that is linked exclusively to RNA polymerase II transcription. Prokaryotic RNA polymerases such as bacteriophage T3, T7 and SP6 polymerases do not  
10 provide messenger RNAs with such a CAP structure, leading to inefficient translation in eukaryotic systems (Fuerst et al. J. Mol. Biol:206:333-348 (1989)).

One way to improve translation of uncapped mRNAs in eukaryotic systems is by the insertion of an  
15 internal ribosome entry site (IRES) sequence 5' of the coding sequence. For example, Elroy-Stein, et al., Proc. Natl. Acad. Sci. USA 87:6743-6747 (1990), describe the cloning of the untranslated region of the ECMV virus downstream of the T7 promoter in order to  
20 enhance the efficiency of translation. In other systems translation of T7-derived transcripts may be enhanced by addition of a CAP structure derived from a capped transcript. For example, in Trypanosoma a 5' CAP structure is added to T7 generated RNA transcripts  
25 by a natural occurring trans-splicing reaction (Wirtz et al. NAR 22:3887-3894 (1994)).

#### Trans-splicing in *C. elegans*.

30 In *C. elegans* many mRNAs contain an identical short leader sequence, designated the spliced leader (SL). This splice leader is donated by a small RNA (SL RNA) via a trans-splicing reaction. This trans  
35 splicing was first observed by Krause et al., Cell 49:753-61 (1987). The splice leader RNA exists as a small nuclear ribonucleoprotein particle and has the trimethylguanosine cap that is characteristic of

- 3 -

eukaryotic small nuclear RNAs. The trimethylguanosine cap present on the spliced leader RNA is transferred to the pre-mRNA during the trans-splicing reaction. Thereafter, the trimethylguanosine cap is maintained on the mature mRNA (Van Doren et al., Mol. Cell. Biol. 10:1769-1772 (1990). The trans-splicing signal for such a splice leader is essentially an intron missing only the 5' splice site, designated an 'outtron'. An outtron has essentially all the intron sequence including a trans-splice acceptor site homologous to a UUUCAG sequence preceded by a AU rich region (Conrad et al., NAR 21:913-919 (1993). Introduction of an outtron into the 5' untranslated region of a *C. elegans* gene converts it to a trans-spliced gene (Conrad et al., EMBO J. 12:1249-1255 (1993); Conrad et al. Mol. Cell Biol. 11:1931-1926 (1991)) and introduction of donor sites in a natural trans-spliced *C. elegans* gene prevents trans-splicing and converts it into a more conventional gene.

20

Description of the invention.

Until recently, expression of heterologous and homologous genes in *C. elegans* was mainly achieved by linking an appropriate coding sequence to a selected *C. elegans* promoter. The present inventors have recently demonstrated that the recombinant gene expression in *C. elegans* can be based on the prokaryotic T7 expression system (WO 00/01846). However, the present inventors found that the expression system was far from being efficient, or at least the resulting expression was much lower than would be expected from this T7 related expression system. It was concluded that this low expression was mainly due to RNA instability or translation arrest. Furthermore, it was reasoned that fundamental differences between prokaryotic and eukaryotic

35

- 4 -

expression systems, particularly the requirement for capping of the 5' end of the mRNA for efficient translation in eukaryotic systems, was the main reason for this unexpectedly low expression.

5       The inventors have now developed a solution to the problem of the inefficiency of the T7 system in eukaryotic host cells and organisms, particularly in *C. elegans*, and have constructed a generally applicable expression system which allows for the  
10       efficient expression of genes, DNA, cDNA, peptides and proteins under the regulation of the T7 promoter in *C. elegans*.

          Therefore, in accordance with a first aspect of the invention there is provided a DNA construct  
15       comprising a bacteriophage promoter operably linked to an outtron sequence.

          It is an essential feature of the DNA construct of the invention that the bacteriophage promoter and the outtron sequence are "operably linked", that is to  
20       say they are arranged in a relationship permitting them to function in their intended manner. In this case, the bacteriophage promoter is positioned upstream of the outtron sequence such that it is capable of promoting transcription of the outtron  
25       sequence upon binding of an appropriate RNA polymerase, with the outtron sequence forming the extreme 5' end of the resulting transcript.

          The DNA construct may further comprise at least one restriction enzyme recognition site positioned  
30       downstream of and proximal to the outtron sequence. Advantageously, the DNA construct may contain multiple restriction sites forming a multi-cloning site. The purpose of the restriction site/multi-cloning site is to facilitate cloning of a heterologous or homologous  
35       DNA fragment downstream of the outtron sequence. A DNA construct comprising a bacteriophage promoter, an outtron sequence and a restriction site/multi-cloning

- 5 -

site may therefore be referred to hereinafter as an 'outtron cloning construct'.

5 In an outtron cloning construct it is advantageous for the restriction site/multi-cloning site to be positioned fairly proximal to the outtron sequence (e.g. within 100bp) such that a heterologous or homologous sequence inserted at this site may be co-transcribed with the outtron sequence on a single mRNA. However, further sequence elements may be interposed  
10 between the outtron sequence and the restriction site/multi-cloning site. For example, the general purpose vector pDW3123 described in the accompanying examples has a synthetic intron A sequence between the outtron sequence and the multi-cloning site.

15 In one preferred embodiment of the invention, the DNA construct is a replicable cloning vector, such as, for example, a plasmid vector. In addition to the bacteriophage promoter, outtron sequence and optional restriction site/multi-cloning site, the vector may  
20 further contain one or more of the general features commonly found in cloning vectors, for example an origin of replication to allow autonomous replication within a host cell and a selective marker, such as an antibiotic resistance gene. Although not essential,  
25 the vector may also contain a poly-adenylation signal to stabilize and process the 3' end of the mRNA transcribed from the bacteriophage promoter. A preferred example is the 3'UTR from the *C. elegans* unc-54 gene, but any other 3'UTR or polyadenylation  
30 signal may be used.

Outtron-containing DNA constructs according to the invention may be easily be constructed from the component sequence elements using standard recombinant techniques well known in the art and described, for  
35 example, in F. M. Ausubel et al. (eds.), *Current Protocols in Molecular Biology*, John Wiley & Sons, Inc. (1994).

- 6 -

Outron sequences for use in the constructs of the invention may be isolated from natural *C. elegans* genes using standard molecular biology techniques. For example, a natural outtron sequence might be  
5 amplified using the polymerase chain reaction or an equivalent amplification technique using *C. elegans* genomic DNA as a template. Alternatively, synthetic outtron sequences may be synthesised, for example, by annealing two complementary single  
10 stranded oligonucleotides, as illustrated in the accompanying examples. Once a DNA fragment comprising the outtron sequence has been obtained it would be a matter of routine to assemble an outtron construct by linking the outtron in the correct orientation relative  
15 to the bacteriophage promoter.

The sequences of the commonly used bacteriophage promoters, e.g. T7, T3 and SP6, are well known in the art and oligonucleotides containing functional phage promoter sequences can be readily synthesised using  
20 standard oligonucleotide synthesis techniques. It would be a matter of routine to insert such a synthetic promoter sequence into, for example, a plasmid vector backbone containing, for example, an origin of replication a selective marker and a  
25 suitable restriction site. Alternatively, one of the many plasmid vectors containing bacteriophage promoter sequences known in the art may be used as the starting point for the construction of a plasmid-based outtron cloning vector. The known vectors generally contain,  
30 in addition to the phage promoter sequence, one or more restriction sites conveniently positioned downstream of the phage promoter and also a bacterial origin of replication and a selective marker. Once the vector backbone is in place the outtron sequence  
35 may simply be inserted in the appropriate position downstream of the bacteriophage promoter.

In a particularly useful embodiment the invention

- 7 -

provides a DNA construct for use in bacteriophage promoter-driven expression of a polypeptide in a eukaryotic host cell or organism. This construct comprises a bacteriophage promoter operably linked to  
5 a DNA sequence such that it is capable of initiating transcription of the DNA sequence upon binding of an appropriate RNA polymerase to the promoter, wherein the aforesaid DNA sequence comprises an outtron sequence and at least one open reading frame  
10 positioned downstream of the outtron sequence.

The open reading frame may be essentially any protein-encoding DNA sequence bounded by start and stop codons. This protein-encoding DNA sequence may include introns, as both trans-splicing and cis-  
15 splicing can occur together.

A DNA construct according to this embodiment of the invention, which may be referred to hereinafter as an 'outtron expression construct', may be derived from an outtron cloning construct by insertion of a  
20 heterologous or homologous protein-encoding DNA fragment into the restriction site/multi-cloning site. It is essential that the heterologous or homologous DNA fragment be inserted downstream of the outtron sequence such that the two sequences may be co-  
25 transcribed, with the outtron sequence forming part of the 5' untranslated region of the resulting mRNA.

The outtron expression construct may advantageously form an expression vector, such as, for example, a plasmid vector. Most preferably, the  
30 expression vector will be one suitable for use in the nematode worm *C. elegans*. In addition to the bacteriophage promoter, outtron sequence and protein-encoding DNA sequence (open reading frame), the expression vector may further contain one or more of  
35 the general features commonly found in expression vectors, for example an origin of replication to allow autonomous replication within a bacterial host cell

- 8 -

and a selective marker, such as an antibiotic resistance gene. The vector may also contain a poly-adenylation signal to stabilize and process the 3' end of the mRNA transcribed from the bacteriophage promoter. A preferred example is the 3'UTR from the *C. elegans* unc-54 gene, but any other 3'UTR or polyadenylation signal may be used. An additional element, such as for example a synthetic intron, may be interposed between the outtron sequence and the open reading frame.

It is important that the open reading frame is positioned downstream of and proximal to the outtron sequence in the expression construct such that (i) the two elements are co-transcribed to form a single mRNA and (ii) the outtron sequence forms part of the 5' untranslated region of the mRNA. If the appropriate splicing machinery and a supply of SL RNAs is provided by the eukaryotic host cell or organism then the uncapped 5' end of the pre-mRNA transcribed from the expression construct will be replaced with a capped splice leader via the trans-splicing reaction. This will greatly increase the efficiency of translation in a eukaryotic host system.

The use of an outtron sequence at the extreme 5' end of the RNA provides a solution to the problem of reduced expression efficiency in eukaryotic systems wherever the type of promoter/polymerase used to drive gene expression leads to the production of uncapped transcripts, provided that the host cell or organism produces the spliced leader RNAs required for the trans-splicing reaction.

Outtron sequences which may be utilised in accordance with the invention include naturally occurring outtron sequences isolated from SL1-specific *C. elegans* genes (Conrad, R. Functional analysis of a *C. elegans* trans-splice acceptor. *Nucleic Acids Res.* 1993, 21(4), pp913-919; Conrad, R. SL1 trans-splicing

- 9 -

specified by AU-rich synthetic RNA inserted at the 5' end of *Caenorhabditis elegans* pre-mRNA. RNA. 1995, 1(2), pp164-170) and also synthetic outtron sequences which are functionally equivalent to the natural *C. elegans* outtron sequences, including variants of naturally occurring *C. elegans* outtrons. The phrase "functionally equivalent" means that the synthetic intron is recognised by the *C. elegans* trans-splicing machinery and can be trans-spliced to a *C. elegans* splice leader RNA, preferably the SL1 splice leader.

Experimental evidence indicates that trans-splicing in *C. elegans* is signalled by an AU-rich intron-like sequence followed by a splice acceptor site (Conrad et al 1993 and 1995). For the purposes of the present application the terms "outtron" or "outtron sequence" should be interpreted as referring to both the AU-rich region from the 5' end of the pre-mRNA to the trans-splice acceptor site and the trans-splice acceptor site itself. In connection with the DNA constructs of the invention, the terms "outtron" and "outtron sequence" refer to features present in the DNA which encodes the pre-mRNA.

The consensus splice acceptor site for trans-splicing of outtrons and the consensus 3' splice acceptor site for cis-splicing of introns are essentially identical (UUUCAG). Moreover, a normally trans-spliced acceptor site can be efficiently cis-spliced when a donor splice site is inserted upstream within the outtron sequence. It is therefore important that the outtron constructs described herein do not contain any potential splice donor sequence upstream of the splice acceptor within the outtron and downstream of the transcription start site such that it will be transcribed in the mRNA encoded by the construct. If such a site were present then there would be a potential for cis-splicing rather than

- 10 -

trans-splicing.

It has also been observed that the overall length of the outtron has an effect on the efficiency of trans-splicing, longer outtrons in general working better than shorter ones (Conrad et al. 1995). Advantageously, the outtron sequences for inclusion into the outtron constructs described herein should be greater than about 50nt in length.

A synthetic outtron containing an AT stretch and a TTTTCAG sequence has been shown to be functional in *C. elegans*. As illustrated in the accompanying Examples, the insertion of an outtron sequence into the 5' untranslated region of GFP reporter construct, downstream of the promoter and upstream of the GFP open reading frame, is required for optimal expression of bacteriophage RNA polymerase transcribed reporter gene mRNA in *C. elegans*.

Suitable bacteriophage promoters which may be used in the DNA constructs according to the invention include T7, T3 and SP6 promoters, with T7 being the most preferred. As discussed above, these bacteriophage promoters have long been known to be useful tools in molecular biology since they can provide simple and strong expression systems dependent only on the binding of the specific or cognate RNA polymerase.

In a still further aspect, the invention provides a method for expressing a recombinant polypeptide in *C. elegans*, which method comprises:

introducing an outtron expression construct, as described above, said construct being an expression vector suitable for use in *C. elegans*, into a *C. elegans* strain which expresses an RNA polymerase specific for the bacteriophage promoter present in said DNA construct in one or more tissues or cell

- 11 -

types.

An outtron expression vector for use in this method may be constructed by inserting DNA encoding the polypeptide of interest into an outtron cloning  
5 vector, as described above. The vector must be one which is suitable for use in *C. elegans*, plasmid-based vectors are the most preferred.

The *C. elegans* worms are preferably transgenic worms carrying a transgene capable of expressing the  
10 RNA polymerase in one or more tissues or cell types. The term "transgene capable of expressing" as used herein means a nucleic acid molecule comprising a nucleotide sequence encoding the polymerase operably linked to a promoter. The promoter may be any  
15 promoter which functions in *C. elegans* and may be general (i.e. active in substantially all tissues and cell types), tissue-specific, cell type-specific, constitutive, inducible etc. Most preferably, the promoter will exhibit tissue or cell type-specificity.  
20 With the use of a tissue or cell type-specific promoter of the appropriate specificity it is possible to control the site of RNA polymerase expression within *C. elegans* and hence control the site of expression of the recombinant polypeptide.

25 Methods for the construction of transgenic *C. elegans* worms are known in the art and are particularly described by Craig Mello and Andrew Fire, *Methods in Cell Biology*, Vol 48, Ed. H.F. Epstein and D.C. Shakes, Academic Press, pages 452-480.

30

In a further aspect the invention provides a kit for use in recombinant expression of a polypeptide in *C. elegans*, the kit comprising an outtron cloning construct, as described above, and optionally a supply  
35 of *C. elegans* nematode worms expressing an RNA polymerase specific for the bacteriophage promoter

- 12 -

present in the said outtron cloning construct in one or more tissues or cell types.

1     The kit might further contain control inserts and control constructs, e.g. a reporter gene inserts and  
5 constructs which could be used to check efficiency of cloning steps and transfection steps, respectively. It might also contain constructs which may be used as selectable markers in the transfection procedure, e.g. a rol 6 plasmid (see below).

10     The invention further provides methods for the construction of transgenic *C. elegans* expressing a recombinant polypeptide in one or more tissues or cell types. One such method comprises introducing an  
15 outtron expression construct, as described above, said construct being an expression vector suitable for use in *C. elegans* comprising an open reading frame encoding the desired recombinant polypeptide, into a *C. elegans* strain which expresses an RNA polymerase specific for the bacteriophage promoter present in  
20 said DNA construct in one or more tissues or cell types, and isolating transgenic *C. elegans* lines which stably express the said polypeptide. The *C. elegans* strain expressing the polymerase is preferably a transgenic strain carrying a transgene capable of  
25 expressing the RNA polymerase in one or more tissues or cell types, as described above. As aforesaid, transgenic *C. elegans* lines can readily be constructed using standard techniques well known in the art.

30     In an alternative approach, the method may comprise introducing into a background *C. elegans* strain (i) an outtron expression construct, as described above, said construct being an expression vector suitable for use in *C. elegans* comprising an open frame encoding the desired recombinant  
35 polypeptide, and (ii) a DNA construct suitable for expression of an RNA polymerase specific for the

- 13 -

bacteriophage promoter present in the outtron expression construct in one or more tissues or cell types of *C. elegans*, and isolating transgenic *C. elegans* lines which stably express the said polypeptide. The second DNA construct may, advantageously, be an expression vector comprising a nucleotide sequence encoding the polymerase operably linked to a promoter having the appropriate tissue or cell type specificity.

10 In carrying out the methods of the invention one may employ standard techniques well known in the art for construction and selection of transgenic *C. elegans* lines. Such techniques are described, for example, in techniques described in Methods in Cell Biology, vol 84; *Caenorhabditis elegans*: modern biological analysis of an organism, ed. Epstein and Shakes, academic press, 1995. Foreign DNA (e.g. plasmid DNA) may be introduced into *C. elegans* using microinjection or ballistic transformation, as described in the applicant's co-pending International patent application No. WO 99/49066. In order to facilitate the selection of transgenic strains a marker plasmid may be co-introduced with the transgenes. A typical example is the plasmid pRF4 (Mello, C. C. et al. EMBO J. 10, 3959-3970 (1991)) which carries the rol-6 gene. *C. elegans* expressing rol-6 can be identified by screening for the roller phenotype. Any other *C. elegans* dominant selectable phenotypic marker, of which there are many known in the art, may be used to facilitate selection of transgenic lines. A useful example is green fluorescent protein (or any of the equivalent autonomous fluorescent proteins known in the art).

30 In a still further aspect the invention provides transgenic *C. elegans* worms which contain an outtron expression construct, as described above, said

- 14 -

construct being an expression vector suitable for use  
in *C. elegans*, and which further express an RNA  
polymerase specific for the bacteriophage promoter  
present in the outtron expression construct in one or  
5 more tissues or cell types.

The present invention will be further understood  
with reference to the following non-limiting Examples,  
together with the accompanying drawings in which:

10

Figure 1 illustrates the construction of a T7-outtron-  
GFP vector. (A) sequence of the synthetic outtron  
produced by annealing oligonucleotides o-GN59 and o-  
GN60. (B) summary of the strategy used to construct  
15 vector pDW3124.

Figure 2 shows plasmid maps for pDW3123 (outtron  
cloning vector) and pDW3124 (outtron expression vector  
for GFP expression).

20

Figure 3 is a plasmid map of pGN148 which contains a  
T7 RNA polymerase coding sequence under the regulation  
of the *C. elegans* SERCA promoter.

25

Figure 4 illustrates the nucleotide sequence of  
pGN148.

30

Figure 5 illustrates the nucleotide sequence of pDW  
3123 annotated to show the positions of the T7  
promoter, outtron, synthetic intron A, multi-cloning  
site and unc-54 3' UTR sequences and also the  
ampicillin resistance gene.

35

Figure 6 illustrates the nucleotide sequence of pDW  
3124 annotated to show the positions of the T7  
promoter, outtron, synthetic intron A, GFP with introns  
and unc-54 3' UTR sequences and also the ampicillin

- 15 -

resistance gene.

Example 1 -Construction of a T7-outtron-GFP containing vector (pDW3124)

5       A SL1 trans-splice acceptor site (outtron) was cloned into a vector downstream of the T7 promoter and upstream of the GFP to be expressed.

10       A synthetic outtron consisting of two partially overlapping oligonucleotides (o-GN59 and o-GN60, see Figure 1) was inserted into a XbaI/XmaI digested T7 promoter GFP construct. Briefly, 25µl o-GN59 and 25µl o-GN60 (100µM) were denatured for 5 minutes at 94°C, annealed for 30 minutes at 68°C then cooled to 4°C. 1µl of XmaI/XbaI digested pDW3120 and 10µl of the  
15       annealed oligos were then ligated using T4 ligase overnight at 16°C, transformed into competent *E. coli* and analysed by restriction digestion and DNA sequencing, all according to standard molecular biology procedures. The resulting vector was  
20       designated pDW3124 (Figures 1 and 2).

      The outtron contains an AU rich sequence followed by a splice-acceptor site as described by Conrad et al, NAR 21:913-919 (1993) (see Figure 1).

25       Example 2-Construction of a T7-Outtron MCS vector

      A general purpose vector was constructed to facilitate expression of other DNA sequences in *C. elegans* under the control of the T7 promoter. This was done by digesting vector pDW3124 with HindII  
30       (position 179) and PvuII (position 1029) (partial digest) and re-ligating the blunt ends, resulting in vector pDW3123 (Figure 2).

35

- 16 -

Example 3-The expression of heterologous genes in *C. elegans* regulated by the T7 promoter requires trans-splicing.

5 Wild-type *C. elegans* nematodes were co-injected with various combinations of the following test plasmids:

- 1) GFP reporter plasmid  
GFP: pDW2020  
10 outtron-GFP: pDW2024  
T7 promoter-GFP: pDW3120  
T7 promoter-outtron-GFP: pDW3124
- 2) T7 polymerase expression plasmid SERCA T7  
15 polymerase: pGN148 together with pRF-4 (rol-6) as marker.

For every co-injection experiment, a total concentration of 200 ng DNA/ $\mu$ l was used (plasmid  
20 concentration was 50 ng/ $\mu$ l and carrier DNA was added up to 200ng/ $\mu$ l). For every co-injection 115 adult worms were injected.

F1 offspring showing the marker rol-6 phenotype  
25 were isolated and then selected for further study. The next generation (F2) of the roller lines were screened for GFP expression in the pharynx, vulva, tail and body wall muscles. These are the tissues in which the bacteriophage T7 RNA polymerase is known to  
30 be expressed when under the control of the *C. elegans* SERCA promoter (as in the construct pGN148)

The results are shown in Table 1 below, which indicates the number of lines expressing GFP vs total number of lines isolated.

35

- 17 -

	1	2	3
A	Construct	no T7-polymerase construct	with T7-polymerase construct (50ng) pGN148
B	GFP (50ng) pDW2020	0/8	2/6*
C	outtron::GFP (50ng) pDW2024	0/11	3/8*
5 D	T7-promoter::GFP (50ng) pDW3120	0/3	0/5
E	T7-promoter::outtron::GFP (50ng) pDW3124	0/7	13/13

\* GFP-expression most probably result of recombination  
10 in the extrachromosomal array

No GFP expression was observed in the experiments  
15 where the T7 RNA polymerase was absent (cells B2, C2, D2, E2).

In the experiments where the T7 RNA polymerase  
expressing vector was co-injected with GFP vectors  
without a T7 promoter, as in the cells B3 and C3, GFP  
20 expression was sometimes observed. This is probably  
due to recombination events in the extrachromosomal  
arrays, resulting in transcription of GFP directly  
from the SERCA promoter.

In the experiments where the T7 promoter-GFP construct  
25 and the SERCA T7 RNA polymerase were co-injected, no  
GFP expression could be observed (cell D3). In  
contrast, all of the lines isolated from the  
experiments where the GFP transcript contained an  
outtron at its 5' site (n=13) expressed GFP (cell E3).  
30 The outtron is a favourable target for SL1  
trans-splicing. Since SL1 RNA molecules contain a 5'

- 18 -

trimethylguanosine CAP structure which is transferred to the mature mRNA this results in improved translation of the RNA and hence better expression of GFP. Without the outtron the T7 RNA polymerase transcripts do not carry a CAP structure at their 5' end, leading to inefficient translation. The results of this experiment illustrate the importance of trans-splicing for efficient expression of heterologous and homologous genes transcribed by prokaryotic polymerases in *C. elegans*.

SEQUENCE LISTING

SEQ ID NO: 1 Oligonucleotide o-GN59  
15 SEQ ID NO: 2 Oligonucleotide 0-GN60  
SEQ ID NO: 3 Plasmid pDW3123  
SEQ ID NO: 4 Plasmid pDW3124  
SEQ ID NO: 5 Plasmid pGN148

Claims:

1. A DNA construct comprising a bacteriophage promoter operably linked to an outtron sequence.  
5
2. A DNA construct as claimed in claim 1 which further comprises at least one restriction enzyme recognition site positioned downstream of and proximal to the outtron sequence.  
10
3. A DNA construct as claimed in claim 2 which comprises a multi-cloning site positioned downstream of and proximal to the outtron sequence.
- 15 4. A DNA construct as claimed in claim 2 or claim 3 which further comprises a DNA fragment inserted at the said restriction site or at a restriction site within the said multi-cloning site.
- 20 5. A DNA construct as claimed in any one of claims 1 to 4 which is a replicable cloning vector.
6. A DNA construct as claimed in any one of claims 1 to 5 wherein the outtron sequence comprises a  
25 3' splice acceptor site having the sequence TTTCAG preceded by an AT-rich region.
7. A DNA construct as claimed in claim 6 wherein the outtron sequence comprises the nucleotide  
30 sequence illustrated in Figure 1A.
8. A DNA construct as claimed in any one of claims 1 to 7 wherein the bacteriophage promoter is the T7, T3 or SP6 promoter.
- 35 9. A DNA construct for use in bacteriophage promoter-driven expression of a polypeptide in a

- 20 -

eukaryotic host cell or organism, which construct comprises a bacteriophage promoter operably linked to a DNA sequence such that it is capable of initiating transcription of said DNA sequence upon binding of the appropriate RNA polymerase to the promoter, wherein the said DNA sequence comprises an outtron sequence and at least one open reading frame positioned downstream of the outtron sequence.

10. A DNA construct as claimed in claim 9 which is an expression vector.

11. A DNA construct as claimed in claim 9 or claim 10 wherein the outtron sequence comprises a 3' splice acceptor site having the sequence TTTCAG preceded by an AT-rich region.

12. A DNA construct as claimed in claim 11 wherein the outtron sequence comprises the nucleotide sequence illustrated in Figure 1A.

13. A DNA construct as claimed in any one of claims 9 to 12 wherein the bacteriophage promoter is the T7, T3 or SP6 promoter.

14. A kit for use in recombinant expression of a polypeptide in *C. elegans*, the kit comprising a DNA construct as claimed in any one of claims 1 to 3, and optionally *C. elegans* worms expressing an RNA polymerase specific for the bacteriophage promoter present in said DNA construct in one or more tissues or cell types.

15. A method for expressing a recombinant polypeptide in *C. elegans* which method comprises: introducing a DNA construct as claimed in any one

- 21 -

of claims 9 to 13, said construct being an expression vector suitable for use in *C. elegans*, into a *C. elegans* strain which expresses an RNA polymerase specific for the bacteriophage promoter present in  
5 said DNA construct in one or more tissues or cell types.

16. A method of generating transgenic *C. elegans* expressing a recombinant polypeptide, which method  
10 comprises:

introducing a DNA construct as claimed in any one of claims 9 to 13 comprising an open reading frame encoding the recombinant polypeptide, said construct being an expression vector suitable for use in *C. elegans*, into a *C. elegans* strain which expresses an  
15 RNA polymerase specific for the bacteriophage promoter present in said DNA construct in one or more tissues or cell types, and

isolating transgenic *C. elegans* lines which  
20 stably express the said polypeptide.

17. A method of generating transgenic *C. elegans* expressing a recombinant polypeptide, which method comprises:

25 introducing into *C. elegans* (i) a first DNA construct as claimed in any one of claims 9 to 13 comprising an open reading frame encoding the recombinant polypeptide, said construct being an expression vector suitable for use in *C. elegans*, and  
30 (ii) a second DNA construct suitable for expression of an RNA polymerase specific for the bacteriophage promoter present in the first DNA construct in one or more tissues or cell types of *C. elegans*, and

isolating transgenic *C. elegans* lines which  
35 stably express the said polypeptide

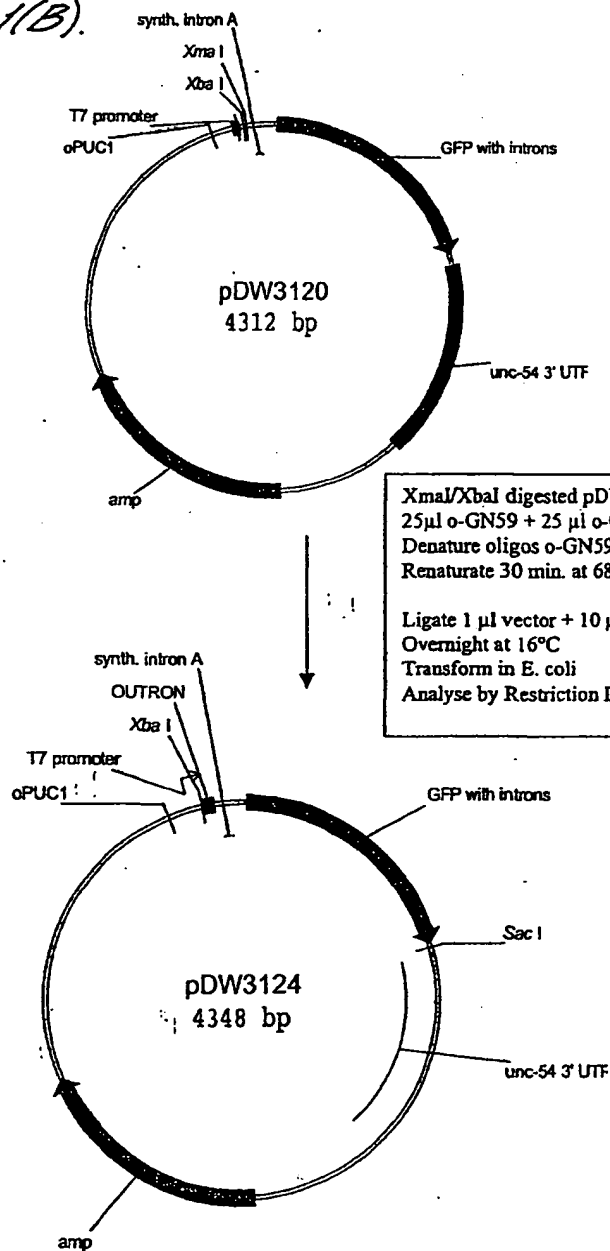
- 22 -

18. Transgenic *C. elegans* which contain a DNA construct as claimed in any one of claims 9 to 13, said construct being an expression vector suitable for use in *C. elegans*, and which further express an RNA  
5 polymerase specific for the bacteriophage promoter present in said DNA construct in one or more tissues or cell types.

FIG. 1(A).

XbaI overhang	SspI	3'splice acceptor	
CTAGATTACAATAATTATACTTATTTGAATATTCAAATTTTCAGAC			o-GN59
TAATGTTGATTAAATATGAATAAACTTATAAGTTTAAAAGTCTGGGCC			o-GN60
		XmaI overhang	

FIG. 1(B).



XmaI/XbaI digested pDW3120  
 25  $\mu$ l o-GN59 + 25  $\mu$ l o-GN60 (100  $\mu$ M)  
 Denature oligos o-GN59 & o-GN60 5 min. at 94°C  
 Renature 30 min. at 68°C, cool to 4°C

Ligate 1  $\mu$ l vector + 10  $\mu$ l oligos with T4 ligase  
 Overnight at 16°C  
 Transform in *E. coli*  
 Analyse by Restriction Digest and sequencing

FIG. 2.

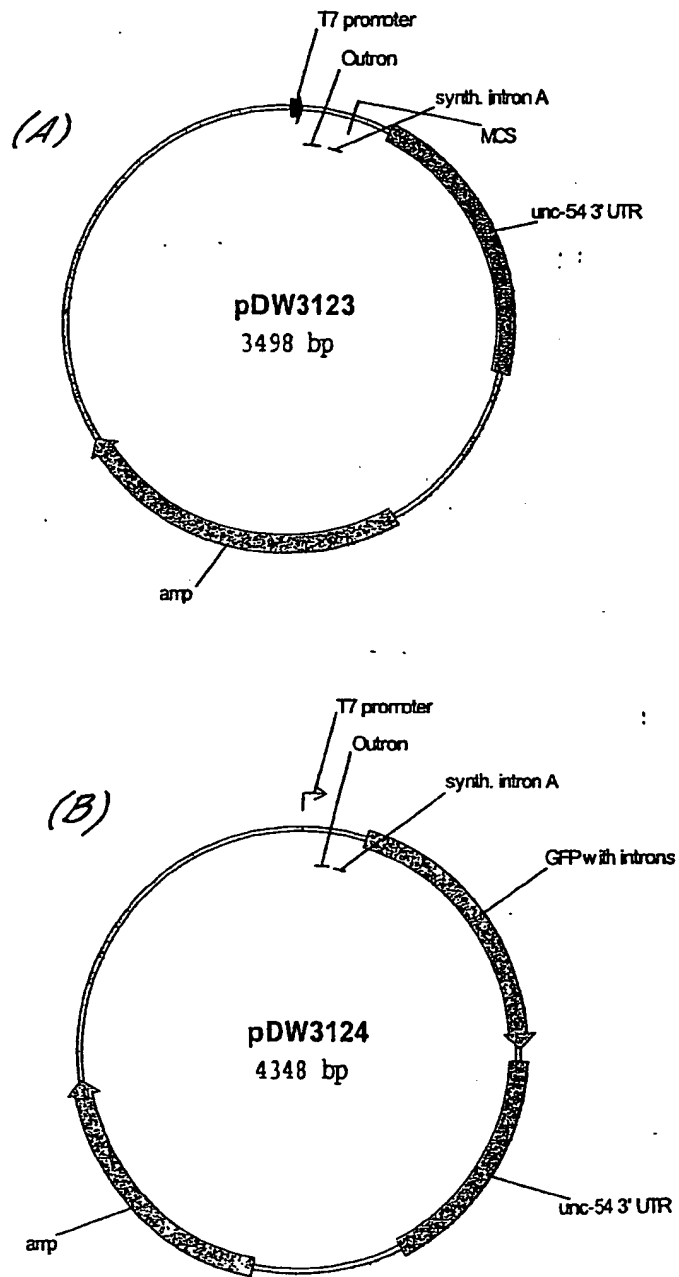
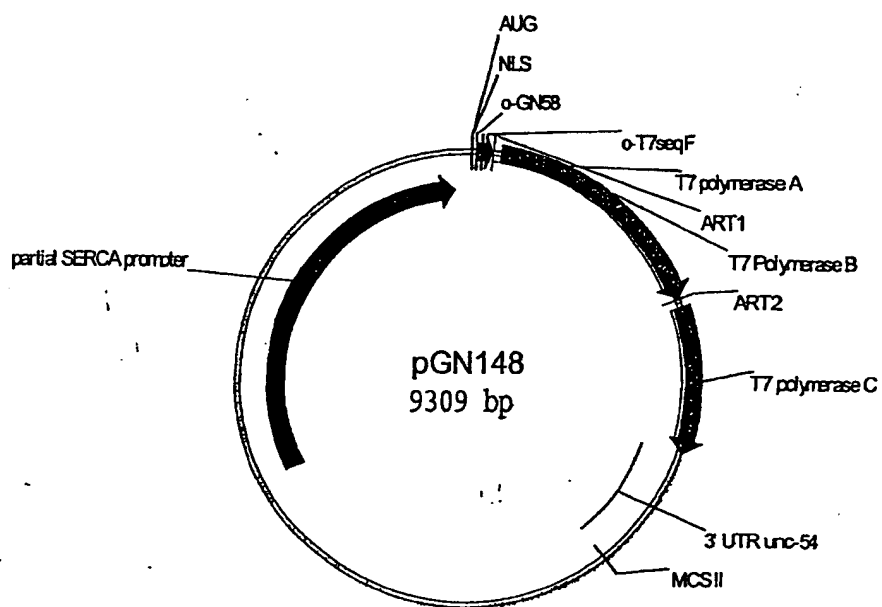


FIG. 3.



*FIG. 4.*

Nucleotide sequence of pGN148

atgactgctccaaagaagaagcgtaaggtaccggtaatgaacacgattaacatcgctaagaacgacttctc  
tgacatcgaaactggctgctatccccgttcaacactctggctgaccattacggtagcggttagctcggttaag  
ttaaacaatctagataactaactaacgattaacatttaaattttcagcgaacagttggcccttgagcatgag  
tcttacgagatgggtgaagcagcgttccgcaagatggttgagcgctcaacttaagctggtaggttgaggga  
taacgctgccgccaagcctctcatcactaccctactccctaagatgattgcacgcatcaacgactggtttg  
aggaagtgaagctaagcgcggcaagcgcggcagagccttcaggttcctgcaagaaatcaagcgggaagcc  
gtagcgtaacatcaccattaaagaccactctggcttgcttaaccagtgctgacaatacaaccgttcaggctgt  
agcaagcgcaatcggtcgggccattgaggacgaggtcgcttcggtcgatccgtgacctgaagctaagc  
acttcaagaaaaacggttaggaacaactcaacaagcgcgtaggggacgctctacaagaaagcatttatgcaa  
gttgctgagggctgacatgctcttaagggctctactcggtggcgaggcggtggtcttcggtggcataaggaaga  
ctctattcatgtaggagtacgctgcacatcgagatgctcattgagtcacacggagtggttagcttacaccggc  
aaaatgctggcgtaggtcaagactctgagactatcgaaactcgacactgaatacgtgaggtctacgca  
accgctgaggtcgctggctggcattcttcggtggttccaaccttcggttagttcctcctaagcggcggtgac  
tggtcattactggtggtggtctatgggttaacggctcgctccttcggtggtggtgctactcacagtaaga  
aagcactgatgctgacgaagcgtttacatgctgaggtgtacaagcgatgaacattgcgcaaaacacc  
gcatggaaaaatcaacaagaaagtccatagcggctcgccaacgtaatacacaagtggaagcattgtccggctcga  
ggacatccctgaggttagcggtgaagaactcccgatgaaacgggaagacatcgacatgaatcctgaggtctc  
tcaccgctggaacgctgctgctgctgctgtgtaccgcaaggacaaggctcgcaagcttcggttatcagc  
cttgagttcatgcttgagcaagccaataagtttgctaaccataaggccatctgggttccttacaacatgga  
ctggcgcggtcggtttacgctgtgtcaatgttcaaccgcgaagctaacgatatgaccaaggactgctta  
cgctggcgaaaggtaaaccaatcggttaagggaaggttactactggctgaaaaatccacgggtgcaaacgtgctg  
ggtgtcgataagggttcctgttccttgagcgcatcaagttcattgaggaaaaccacgagaacatcatggcttg  
cgctaagtctccactggagaacacttggtgggctgagcaagatttcctgcttccttcggttcctgct  
ttgagtagcgtgggttacagcaccacggcctgagctataactgttccttcctggtggcgtttgacgggtct  
tgctctggcatccagcaacttctcgcgatgctccgagatgaggtagggtggtcgcgcggttgtaagtttaaa  
ctctatcctactaactaacgaagcttatttaaattttcagaacttgcttcctagtgaacacgttcaggaca  
tctacgggattgttgctaagaaagtcaacgagattctacaagcagacgcaatcaatgggacgataacgaa  
gtagttaccgtgaccgatgagaacactggtgaaatctctgagaaagtcaagctgggcaactaaggcactggc  
tggtaaatggctgggttaccgtgttactcgagtggtgactaagcgttcagtcacgacgtggcttacgggt  
ccaaagagcttcggcttcctgcaacagtgctgggaagataccattcagccagctattgatccggcaagggt  
ctgatgttactcagccgaatcaggctgctggatcatggctaaagctgatctgggaatctgtgagcgtgac  
ggtggtgagctcggttgaaagcaatgaactggcttaagctgctgctaagctgctggctgctgaggtcaag  
ataagagcactggagagattcttcgcaagcgttcgctgtgcatgggtcactccggatgggttcctctgtg  
tggcaggaataacaagaagcctattcaaacgcggttgaaactgatgttcctcggtcagttccgcttacagcc  
taccattaacaccaacaagatagcgagattgatgcacacaaacaggagctcgggtatcgctcctaacttg  
tacacagccaagacggttagccaccttcgttaagactgtagtgtgggcacacgagaagtacggaatcgaatct  
tttgactgattcacgactccttcgggtaccattccggctgacgtcgcaacctgttcaaagcagtgccgca  
aactatgggtgacacatatgagtccttgtagtgactggctgatttctacgaccagttcgctgaccagtgc  
acgagctcctaattggacaaaatgccagcacttcgggttaaggttaacttgaacctccgtgacatcttagag  
tcggacttcgcttcgctgaagaattccaactgagcgcgggtcgctaccattaccaactctgctgggtgca  
aaaataatagggcgctgtcatcagagtaagtttaaactgagttctactaactaacgagtaataatttaa  
ttttcagcatctcgcccgctgctctgacttctaagtcctaattactctcaacatccctacatgctcttt  
ctccctgtgctccccccttattttgttattatcaaaaaaacttcttcttaatttctttgttttttagc  
ttcttttaagtcacctctaacaatgaaattgtgtagattcaaaaaatagaattaattcgtaataaaaagtcg  
aaaaaaattgtctccctcccccatataataattctatcccaaatctacacaatgttctgtgtacac  
ttcttatgttttttacttctgataaatttttttgaaacatcatagaaaaaacgcacacaaaatacct  
tatcatatgttacgtttcagtttatgaccgcaatttttatttcttcgacgcttggtggtctctcatgacgtc  
aatcatgctcatcgtaaaaaagttttggagttttttggaatttttcaatcaagtgaagtttatgaaat  
taatttctctgcttttcttttgggggtttccctattgtttgtcaagagtttcgaggacggcgcttttc  
ttgctaaaatcacaagttatgatgagcagatgaagaagatcggaagaaggtttgggtttgaggtcag  
tggagggtgagtagaagttgataatttgaaagtgagtagtgctatgggggtttttgccttaaatgacaga  
atacattcccaatatacacaataactgtttctactagtcggccgtacggggccctttcgtctcgcgct  
ttcggtgatgacggtgaaaacctctgacacatgcagctcccgagacggtcacagcttgctgtgaagcgga  
tgccgggagcagacaagcccgctcagggcgctcagcggtgttggggggtgctgggggtggcttaactatg  
cggcatcagagcagattgtactgagagtgacccatgcgggtgtgaaataccgcacagatgcgttaaggaga

*FIG. 4 (CONTINUED 1.)*

aaataccgcatcaggcgcccttaaggccctcgatgacgcctatttttataggttaatgtcatgataataa  
tgggtttcttagacgtcaggtggcacttttcggggaaatgtgcgaggaaacccctatttggttatttttctaa  
atacattcaaatatgtatccgctcatgagacaataaccctgataaatgcttcaataatattgaaaaaggaa  
gagtatgagtatcaacattttccgtgtcgcccttattcccttttttcgggcattttgccctcctgtttttg  
ctcaccagaaacgctgggtgaaagttaaagatgctgaagatcagttgggtgcacgagtggtttacatcgaa  
ctggatctcaacagcggttaagatccttgagagttttcgcccgaaagacgttttccaatgatgagcacttt  
taaagtctgtctatgtggcgcggtattatcccgatttgacgcccgggcaagagcaactcgggtcgccgcatac  
actattctcagaatgacttgggtgagtgactcaccagtcacagaaaagcatcttacggatggcatgacagta  
agagaattatgagtgctgccataaccatgagtgataaacactgcccgaacttacttctgacaacgatcgg  
aggaccgaaggagctaacggcttttttgcaacaatggtgggatcatgtaactcgccctgacgttgggaac  
cggagctgaatgaagccataccaaacgacgagcgtgacaccagatgcctgtagcaatggcaacaacgttg  
cgcaaacatttaactggcgaaacttacttacttagcttcccggcaacaattaatagactggatggaggcgga  
taaagtgcaggaccacttctgcgtcgcccttccggctggctgggtttattgctgataaatctggagccg  
gtgagcgtgggtctcgcggtatcttgacgactggggccagatggttaagccctcccgatcgtagttatc  
tacacgacggggagtcaggcaactatggatgaacgaaatagacagatcgctgagataggtgcctcactgat  
taagcattggtaactgtcagaccaagtttactcatatatacttttagattgatttaaaacttctattttta  
ttaaaggatctaggtgaagatccttttgataatctcatgacaaaaatcccttaacgtgagttttcgttc  
cactgagcgtcagacccttagaaaaagatcaaaggatccttcttgagatcctttttctgcgctgaatctg  
ctgcttgcaaaaaaaaccccgctaccagcggtgggtttgtttgcggatcaagagctaccaactcttt  
ttccgaaggtaactggcttcagcagagcgcagataccaaatctgtccttctagtgtagccgtagttaggc  
caccacttcaagaactctgtagcaccgctacatacctcgctctgtaactctgttaccagtggtgctgc  
cagtgccgataagtctgtcttaccgggttggaactcaagacgatagttaccggataaggcgagcggtcgg  
gctgaacggggggtctgtgcacacagcccagcttggagcgaaacgacctacaccgaactgagatacctacag  
cgtgagcattgagaaagcggcagcttcccgaaggagaaaggcgagcaggtatccggtaagcggcaggggt  
cggaacaggagagcgcagggagcttccagggggaaacgcctggatctttatagtcctgtcgggtttc  
gccacctctgacttgagcgtcgatttttgatgctcgtcagggggggcgagcctatggaaaaacggcagc  
aacgcggcctttttcgaagagaaacttttgagtgcccttttgctcacatgtcttctcgtgcttctccc  
tgattctgtggataaccgtattaccgccttgagtgagctgataccgctcgccgcagccgaacgaccgagc  
gcagcgagtcagtgagcaggaagcgggaagagcggcacaatcgcaaacccgctctccccgcgcttggccg  
attcattaatgagctggcagcagaggtttcccgactggaaagcgggagtgagcgaacgcaatattatgt  
gagtgagctcactcatttaggcaccccgctttacactttatgcttccggctcgatggtgtgtggaaattg  
tgagcggataacaatttcacacaggaaacagctatgacctgattacgccaagctgtaagttaaacaatga  
tcttactaactaactattctcatttaaattttcagagcttaaaaaatgggtgaaatcactcacaacgatgga  
tacgctaacaacttggaatgaaataagcttgcatgctgcagagcaaaaaaactgcttttctctgcaa  
aatcggcgtcttttcaagagaaacttttgagtcggcgagcatttcttctctgacttctctcttt  
ccgcaaaaagcctagcattttttatgataatttgattacacacactcagagttcttcgacatgataaagt  
gtttcattggcactcgccctaacagtcacatgacaagggcggtatttatcgatcgatattgaagacaaact  
ccaaatgtgtgctcattttggagccccgttggggcgagctgctctcaatatattactagggagacgaggag  
ggggaccttatcgaaactgcagtgagcattcttcttcttcttgactcttctcactctctcacacatta  
atcgattcattagactcccatattccttgatgaaggtgtgggttttttagcttttttcccgatttgtaaaag  
gaagaggctgacgatgttaggaaaaagagaacggagccgaaaaaacatccgtagtaagtcttcttttaag  
ccgacactttttagacagcattcgcgctagttttgaagtttaaaattttaaaaaataaaaaatagtttcaa  
tttttttaattactaaaataggcaaaagtttttcaagaactctagaaaaactagcttaattcatgggtac  
tagaaaaatcttgtttttaaatttaatttatttcttaagatgtaattacgagaagcttttttgaaaaatct  
caatataaagaatttgccgatttagaataaaagtcttcagaaatgagtaaaagctcaaattagaagtttgt  
ttttaaggaaaaaacacgaaaaaagaacactatttatctttctccccgcgtaaaaatagttgtgtgat  
aatagtgatccgctgtctatttgactcggtcttccacacgctgcttctctcacttgacccaacaggaaaa  
aaaaaacatcagcttgagacgggtgaattgccttatcaagagcgtcgctcttccaccagtaacaaaaaa  
aatttgggtttcttactttatattttaggtcacaaaaaaaagtgtatgcagttttgtgggtcggttgt  
ctccacaccacctccgctccagcagcacacaatcatctctgtgtgttctcgacgatctcttgatgcccgc  
ggtcgatgaatgaccacattcgacgcgcaactacacaccacactcactttcggtgtgtattactacacgtca  
tcgttgttcgtagtctccgctcttctgctcccaactcactctctatttcccttgggtgtattgattttt  
tttaaatgggtacaccactctcgaggttctacettctgttttccgtccatttagatttctatctgaaatt  
tttttaaaattttaggccagagagttctagttctgtttctaaagtctaggtcagacatacatttctatt  
tctcatcaaaaaaaagttagataaagaaaactggttattcagaaagagtg-gtctcggtgaaattgattca  
aaaaaaatccccacccctcgcttggttctcaaaatgatgatcaacgga-tttttcttctcgattcaat  
ttttgtcgtcgctctgtctgccaaagtgtgtgtgtccgagcaaaaagatgagagaatttacaacagaaatg

aaaaaaagtgtggccaaataatgaagttttatccgagattgatgggaaagatatataatgttctttacggttt  
 ggaggggagagagagatagattttcgcatacaactccgccttttacatgtcttttagaatctaaaatagat  
 ttttctcatcatttttaatatagaaaaatcgagaaattacagtaattttcgcaattttcttgccaaaaatacacg  
 aaatttgtgggtctcgccacagatctcggtcttagtgggtcatttgggtttaaaagtttataaaatttcaaa  
 tctagtgttttaattttccgcataaattggacctaaaatgggtttttgtcatcattttcaacaagaataatcgta  
 aaatcctgttgttttcgcaattttcttttcaaaaaatacacgaaatatatggtaatttcccgaaatatttgagg  
 gtctcgccacgatttccagtccagtgggcaggatttatcacgaaaaaagttcgcttagtctcacatttccg  
 gaaaaaccgaatctaaattagttttttgtcatcattttgaacaaaaaatcgagacatccctatagtttcgca  
 attttcgtcgctttttctccaaaaatgacagttcagaataaaaattcgtcggaatccggaccatgatct  
 tttctccccgttttttcttttattttttttatctacatggattgactaaaggtcacccaccgccgagtgtg  
 tgccatatcacacacacacacacacacacaaatgtcgagattttatgtgttttccctgcttgatttccgttccgt  
 tgtctctctctctctattcatcttttgagccgagaagctccagagaatggagcacacaggatcccgccgcg  
 cgtatgctcgtcgggagatggcgcgcgcctgggaagccgcgcgagatatcagggaagatcgtctgatttctcc  
 tcggatgcaccctcatctctcagttttctccgctgttactccctgcgaacctgatatttccccgttgtc  
 taagaagatgttttttatttttactttacaccgggtctctctctctcgcagcacagctcagttgtgctgtg  
 tgtcgggtcctgccaccggcgccctcatct  
 ctattcattcttattccttttcatcatcaaaactagcatttcttactttatttttttttcaattttcaat  
 tttcagataaaaacaaactacttgggttacagccgtcaacagatccccgggattggccaaaggaccacaaag  
 gtagtgttcgaatgatactaacataacatagaacattttcaggaggacccttgcttggagggtaccgagct  
 cagaaaaa

*FIG. 5.*

```

=====
T7 promoter                                     Outron
=====
1  AGCTTGGCGC CTAATACGAC TCACTATAGG GCTGCAGGTC GACTCTAGAT TACAACTAAT TATACTTATT
   TCGAACCGCG GATTATGCTG AGTGATATCC CGACGTCCAG CTGAGATCTA ATGTTGATTA ATATGAATAA

Outron                                     synth. intron A
=====
71  TGAATATTCA AATTTTCAGA CCCGGGATTG GCCAAAGGAC CCAAAGGTAT GTTTCGAATG ATACTAACAT
    ACTTATAAGT TAAAAGTCT GGGCCCTAAC CGGTTTCCTG GGTTTCATA CAAAGCTTAC TATGATTGTA

synth. intron A                               MCS
=====
141 AACATAGAAC ATTTTCAGGA GGACCTTGG CTAGCGTCCT GCTGGGATTA CACATGGCAT GGATGAACTA
    TTGTATCTTG TAAAAGTCTC CCTGGGAACC GATCGCAGGA CGACCCTAAT GTGTACCGTA CCTACTTGAT

                                     unc-54 3' UTR
=====
211 TACAAATAGG GCCGGCCGAG CTCCGCATCG GCCGCTGTCA TCAGATCGCC ATCTCGCGCC CGTGCCCTCTG
    ATGTTTATCC CGGCCGGCTC GAGGCGTAGC CGGCACAGT AGTCTAGCGG TAGAGCGCGG GCACGGAGAC

                                     unc-54 3' UTR
=====
281 ACTTCTAAGT CCAATTACTC TTCAACATCC CTACATGCTC TTCTCCCTG TGCTCCACC CCTATTTTT
    TGAAGATTCA GGTTAATGAG AAGTTGTAGG GATGTACGAG AAAGAGGGAC ACGAGGGTGG GGGATAAAAA

                                     unc-54 3' UTR
=====
351 GTTATTATCA AAAAAGCTT TTCTTAATTT CTTTGTCTTT TAGCTTCTTT TAAGTCACCT CTAACAATGA
    CAATAATAGT TTTTGTGAAG AAGAATTAA GAACAAAAA ATCGAAGAAA ATTCAGTGGG GATTGTTACT

                                     unc-54 3' UTR
=====
421 AATTGTGTAG ATTCAAAAAT AGAATTAAAT CGTAATAAAA AGTCGAAAAA AATTGTGCTC CCTCCCCCA
    TTAACACATC TAAGTTTITA TCTTAATTAA GCATTATTTT TCAGCTTTT TTAACACGAG GGAGGGGGGT

                                     unc-54 3' UTR
=====
491 TTAATAATAA TTCATCCCA AAATCTACAC AATGTTCTGT GTACACTTCT TATGTTTTT TTACTTCTGA
    AATTATTATT AAGATAGGCT TTAGATGTG TTACAAGACA CATGTGAAGA ATACAAAAA AATGAAGACT

                                     unc-54 3' UTR
=====
561 TAAATTTTTT TTGAACATC ATAGAAAAA CCGCACACAA AATACCTTAT CATATGTTAC GTTTCAGTTT
    ATTTAAAAA AACTTTGTAG TATCTTTTTT GCGTGTGTT TTATGGAATA GTATACAATG CAAAGTCAAA

unc-54 3' UTR
=====
631 ATGACCGCAA TTTTATTTC TTCGCACGTC TGGGCTCTC ATGACGTCAA ATCATGCTCA TCGTGAAAAA
    TACTGGCGTT AAAAATAAAG AAGCGTGCAG ACCCGGAGAG TACTGCAGTT TAGTACGAGT AGCACTTTTT

unc-54 3' UTR
=====
701 GTTTTGGAGT ATTTTGGAA TTTTCAATC AAGTGAAAGT TTATGAAATT AATTTCCCTG CTTTGCTTT
    CAAAACCTCA TAAAACCTT AAAAAGTTAG TTCACTTTCA AATACTTTAA TAAAAGGAC GAAAACGAAA

unc-54 3' UTR
=====
771 TTGGGGGTTT CCCCTATTGT TTGTCAAGAG TTTGAGGAC GCGTTTTTTC TTGCTAAAAT CACAAGTATT
    AACCCCAAAA GGGGATAACA AACAGTTCTC AAAGCTCCTG CCGCAAAAAG AACGATTTTA GTGTTTCATA

unc-54 3' UTR
=====
841 GATGAGCAGG ATGCAAGAAA GATCGGAAGA AGGTTTGGGT TTGAGGCTCA GTGGAAGGTG AGTAGAAGTT
    CTACTCGTGC TACGTTCTTT CTAGCCTTCT TCCAAACCCA AACTCCGAGT CACCTCCAC TCATCTTCAA

unc-54 3' UTR
=====
911 GATAATTTGA AAGTGGAGTA GTGTCTATGG GGTTTTGCC TTAATGACA GAATACATTC CCAATATACC
    CTATTAACT TTCACCTCAT CACAGATACC CAAAAACGG AATTACTGT CTTATGTAAG GGTATATGG

unc-54 3' UTR
=====
981 AAACATAACT GTTTCCTACT AGTCGGCGT ACGGGCCCTT TCGTCTCGCG CGTTTCGGTG ATGACGGTGA
    TTTGTATTGA CAAAGGATGA TCAGCCGGCA TGCCCGGAA AGCAGAGCGC GCAAAGCCAC TACTGCCACT

```

## FIG. 5 (CONTINUED 1.)

1051 AAACCTCTGA CACATGCAGC TCCCGGAGAC GGTACAGCT TGTCTGTAAG CGGATGCCGG GAGCAGACAA  
TTTGGAGACT GTGTACGTCG AGGGCCTCTG CCAGTGTCTGA ACAGACATTC GCCTACGGCC CTCGTCTGTT

1121 GCCCGTCAGG GCGCGTCAGC GGGTGTGGC GGGTGTCCGG GCTGGCTTAA CTATGCCGCA TCAGAGCAGA  
CGGGCAGTCC CGCGCAGTCG CCCACAACCG CCCACAGCCC CGACCGAATT GATACGCCGT AGTCTCGTCT

1191 TTGTACTGAG AGTGCACCAT ATGCGGTGTG AAATACCGCA CAGATGCCGA AGGAGAAAAT ACCGCATCAG  
AACATGACTC TCACGTGGTA TACGCCACAC TTTATGGCGT GTCTACGCAT TCCTCTTTTA TGGCGTAGTC

1261 GCGGCCTTAA GGGCCTCGTG ATACGCCTAT TTTTATAGGT TAATGTCATG ATAATAATGG TTTCTTAGAC  
CGCCGGAATT CCCGGAGCAC TATGCGGATA AAAATATCCA ATTACAGTAC TATTATTACC AAAGAATCTG

1331 GTCAGGTGGC ACTTTTCGGG GAAATGTGCG CGGAACCCCT ATTTGTTTAT TTTTCTAAAT ACATTCAAAT  
CAGTCCACCG TGAAGAGCCC CTTTACACGC GCCTTGGGGA TAAACAAATA AAAAGATTTA TGTAAGTTTA

amp

1401 ATGTATCCGC TCATGAGACA ATAACCCTGA TAAATGCTTC AATAATATTG AAAAAGGAAG AGTATGAGTA  
TACATAGGCG AGTACTCTGT TATTGGGACT ATTTACGAAG TTATTATAAC TTTTTCCTTC TCATACTCAT

amp

1471 TTCAACATTT CCGTGTCCGC CTATTCCCT TTTTGGCGG ATTTGCGCTT CCTGTTTTG CTCACCCAGA  
AAGTTGTAAA GGCACAGCGG GAATAGGGA AAAACGCGG TAAACGGAA GGACAAAAAC GAGTGGGTCT

amp

1541 AACGCTGGTG AAAGTAAAG ATGCTGAAGA TCAGTTGGGT GCACGAGTGG GTTACATCGA ACTGGATCTC  
TTGCGACCAC TTTCAATTTT TACGACTTCT AGTCAACCCA CGTGCTCACC CAATGTAGCT TGACCTAGAG

amp

1611 AACAGCGGTA AGATCCTTGA GAGTTTCGC CCCGAAGAAC GTTTCCCAAT GATGAGCACT TTTAAAGTTC  
TTGTGCCCAT TCTAGGAAT CTCAAAGCG GGGCTTCTTG CAAAAGGTTA CTACTCGTGA AAATTTCAAG

amp

1681 TGCTATGTGG CGCGGTATTA TCCCGTATTG ACGCCGGGCA AGAGCAACTC GGTGCGCGCA TACACTATTC  
ACGATACACC GCGCCATAAT AGGGCATAAC TGCGGCCCGT TCTCGTTGAG CCAGCGGCGT ATGTGATAAG

amp

1751 TCAGAAATGAC TTGGTTGAGT ACTCACCAGT CACAGAAAAG CATCTTACGG ATGGCATGAC AGTAAGAGAA  
AGTCTTACTG AACCAACTCA TGAGTGGTCA GTGTCTTTT GTAGAATGCC TACCGTACTG TCATTCTCTT

amp

1821 TTATGCAGTG CTGCCATAAC CATGAGTGAT AACACTGCGG CCAACTTACT TCTGACAACG ATCGGAGGAC  
AATACGTCAC GACGGTATTG GTACTACTA TTGTGACGCC GGTGAATGA AGACTGTTGC TAGCCTCCTG

amp

1891 CGAAGGAGCT AACCGCTTTT TTGCACAACA TGGGGATCA TGTAACTCGC CTTGATCGTT GGGAAACGGG  
GCTTCTCGA TTGGCGAAAA AACGTGTTGT ACCCCCTAGT ACATTGAGCG GAACTAGCAA CCCTTGGCCCT

amp

1961 GCTGAATGAA GCCATACCAA ACGACGAGCG TGACACCACG ATGCCTGTAG CAATGGCAAC AAGCTTGGCG  
CGACTTACTT CGGTATGGTT TGCTGCTCGC ACTGTGGTGC TACGGACATC GTTACCGTTG TTGCAACGCG

*FIG. 5 (CONTINUED 2.)*

amp  
=====

2031 AAAGTATTAA CTGGCGAACT ACTTACTCTA GCTTCCCGGC AACAAATTAAT AGACTGGATG GAGGCGGATA  
TTTGATAATT GACCGCTTGA TGAATGAGAT CGAAGGGCCG TTGTTAATTA TCTGACCTAC CTCCGCCTAT

amp  
=====

2101 AAGTTGCAGG ACCACTTCTG CGCTCGGCCC TTCCGGCTGG CTGGTTTATT GCTGATAAAT CTGGAGCCGG  
TTCAACGTCC TGGTGAAGAC GCGAGCCGGG AAGGCCGACC GACCAAATAA CGACTATTTA GACCTCGGCC

amp  
=====

2171 TGAGCGTGGG TCTCGCGGTA TCATTGCAGC ACTGGGGCCA GATGGTAAGC CCTCCCGTAT CGTAGTTATC  
ACTCGCACCC AGAGCGCCAT AGTAACGTGC TGACCCCGGT CTACCATTCCG GGAGGGCATA GCATCAATAG

amp  
=====

2241 TACACGACGG GGAGTCAGGC AACTATGGAT GAACGAAATA GACAGATCGC TGAGATAGGT GCCTCACTGA  
ATGTGCTGCC CCTCAGTCCG TTGATACCTA CTTGCTTTAT CTGTCTAGCG ACTCTATCCA CGGAGTGACT

amp  
=====

2311 TTAAGCATTG GTAAGTGTCA GACCAAGTTT ACTCATATAT ACTTTAGATT GATTTAAAC TTCATTTTTA  
AATTCGTAA CATTGACAGT CTGGTTCAAA TGAGTATATA TGAAATCTAA CTAAATTTTG AAGTAAAT

2381 ATTTAAAGG ATCTAGGTGA AGATCCTTTT TGATAATCTC ATGACCAAAA TCCCTTAACG TGAGTTTTCG  
TAAATTTTCC TAGATCCACT TCTAGGAAAA ACTATTAGAG TACTGGTTTT AGGGAATTGC ACTCAAAAGC

2451 TTCCACTGAG CGTCAGACCC CGTAGAAAAG ATCAAAGGAT CTTCTTGAGA TCCTTTTTTT CTGCGCGTAA  
AAGGTGACTC GCAGTCTGGG GCATCTTTTC TAGTTTCCTA GAAGAACTCT AGGAAAAAAA GACGCGCATT

2521 TCTGCTGCTT GCACACAAAA AAACCACCGC TACCAGCGGT GGTTTGTTTG CCGGATCAAG AGCTACCAAC  
AGACGACGAA CGTTTGTTTT TTTGGTGCGC ATGCTCGCCA CCAACAAAC GGCCTAGTTC TCGATGGTTG

2591 TCTTTTTCCG AAGGTAACG GCTTCAGCAG AGCGCAGATA CCAAATACTG TCCTTCTAGT GTAGCCGTAG  
AGAAAAAGGC TTCCATTGAC CGAAGTCGTC TCSCGTCTAT GGTTTATGAC AGGAAGATCA CATCGGCATC

2661 TTAGGCCACC ACTTCAAGAA CTCTGTAGCA CCGCCTACAT ACCTCGCTCT GCTAATCCTG TTACCAGTGG  
AATCCGGTGG TGAAGTCTT GAGCATCGT GGCAGATGTA TGGAGCGAGA CGATTAGGAC AATGCTCACC

2731 CTGCTGCCAG TGGCGATAAG TCGTGCTTAA CCGGGTTGGA CTCAAGACGA TAGTTACCGG ATAAGGCGCA  
GACGACGGTC ACCGCTATTC AGCACAGAAT GGCCTAACCT GAGTCTGCT ATCAATGGCC TATTCGCGT

2801 GCGGTGCGGC TGAACGGGGG GTTCGTGCAC ACAGCCCAGC TTGGAGCGAA CGACCTACAC CGAAGTGA  
CGCCAGCCCG ACTTGCCCC CAAGCACGTG TGTGCGGTG AACCTCGCTT GCTGGATGTG GCTTGACTCT

2871 TACCTACAGC GTGAGCATTG AGAAAGCGCC ACGCTTCCCG AAGGGAGAAA GGCGGACAGG TATCCGGTAA  
ATGGATGTG CACTCGTAAC TCTTCGCGG TGCGAAGGGC TTCCCTCTT CCGCCTGTCC ATAGGCCATT

2941 GCGGCAGGGT CGGAACAGGA GAGCGCACGA GGGAGCTTCC AGGGGGAAC GCCTGGTATC TTTATAGTCC  
CGCCGTCCCA GCCTTGCTCT CTGCGTGCT CCCTCGAAGG TCCCCCTTTG CGGACCATAG AAATATCAGG

3011 TGTCGGGTTT CGCCACCTCT GACTTGAGCG TCGATTTTTG TGATGCTCGT CAGGGGGGCG GAGCCTATGG  
ACAGCCGAAA GCGGTGGAGA CTGAATCGC AGCTAAAAAC ACTACGAGCA GTCCCCCGC CTCGGATACC

3081 AAAACGCCA GCAACGCGGC CTTTTACGG TTCTGGCCT TTTGCTGGC TTTTGCTCAC ATGTTCTTTC  
TTTTTGCGGT CGTTGCGCCG GAAAAATGCC AAGGACCGGA AAACGACCGG AAAACGAGTG TACAAGAAAG

3151 CTGCGTTATC CCCTGATTCT GTGGATAACC GTATTACCGC CTTTGAGTGA GCTGATACCG CTCGCCGCGC  
GACGCAATAG GGGACTAAGA CACCTATTGG CATAATGGCG GAACTCACT CGACTATGGC GAGCGCGCTC

3221 CCGAACGACC GAGCGCAGCG AGTCAGTGAG CGAGGAAGCG GAAGAGCGCC CAATACGCAA ACCGCCTCTC  
GGCTTGCTGG CTCGCTCGC TCAGTCACTC GTCCTTCGC CTTCTCGCGG GTTATGCGTT TGGCGGAGAG

3291 CCCGCGGCTT GCGCGATTCA TTAATGCAGC TGGCAGACA GGTTCGCCA CTGGAAAGCG GGCAGTGAGC  
GGCGCGCAA CCGCTAAGT AATTACGTCG ACCGTGCTGT CCAAGGGCT GACCTTTCGC CCGTCACTCG

3361 GCAACGCAAT TAATGTGAGT TAGCTCACTC ATTAGGCACC CCAGGCTTTA CACTTTATGC TTCCGGCTCG  
CGTTGCGTTA ATTACACTCA ATCGAGTGAG TAATCCGTGG GGTCCGAAT GTGAAATACG AAGGCCGAGC

3431 TATGTTGTGT GGAATTGTGA GCGGATAACA ATTCACACA GGAACAGCT ATGACCATGA TTACGCCA  
ATACAACACA CCTAACACT CGCCTATTGT TAAAGTGTGT CCTTGTCGA TACTGGTACT AATGCGGT

*FIG. 6.*

```

=====
T7 promoter                                     Outron
=====
1  AGCTTGGCGC CTAATACGAC TCACTATAGG GCTGCAGGTC GACTCTAGAT TACAACATAAT TATACTTATT
   TCGAACCGCG GATTATGCTG AGTGATATCC CGACGTCCAG CTGAGATCTA ATGTTGATTA ATATGAATAA

Outron                                     synth. intron A
=====
71  TGAATATTCA AATTTTCAGA CCCGGGATTG GCCAAAGGAC CCAAAGGTAT GTTTCGAATG ATACTAACAT
    ACTTATAAGT TAAAAGTCT GGGCCCTAAC CGGTTTCCTG GGTTCCTATA CAAAGCTTAC TATGATTGTA

synth. intron A                                     GFP with introns
=====
141 AACATAGAAC ATTTTCAGGA GGACCCTTGG CTAGCGTCCA CGGTACCATG GGGCGCGCCA TGAGTAAAGG
    TTGTATCTTG TAAAAGTCTT CCTGGGAACC GATCGCAGCT GCCATGGTAC CCCGCGCGGT ACTCATTTC

GFP with introns
=====
211 AGAAGAACTT TTCCTGGAG TTGTCCCAAT TCTTGTGAA TTAGATGGTG ATGTTAATGG GCACAAATTT
    TCTTCTTGAA AAGTGACCTC AACAGGGTTA AGAACAACTT AATCTACCAC TACAATTACC CGTGTTTAA

GFP with introns
=====
281 TCTGTCACTG GAGAGGGTGA AGGTGATGCA ACATACGGAA AACTTACCCT TAAATTTATT TGCCTACTG
    AGACAGTCAC CTCTCCCACT TCCACTACGT TGTATGCCTT TTGAATGGGA ATTTAAATAA ACGTGATGAC

GFP with introns
=====
351 GAAACTACCC TGTTCATGG GTAAGTTTAA ACATATATAT ACTAACAAC CCTGATTATT TAAATTTTCA
    CTTTGTATGG ACAAGGTACC CATTCAAAT TGTATATATA TGATTGATTG GGACTAATAA ATTTAAAGT

GFP with introns
=====
421 GCCAACACTT GTCCTACTT TCTGTATGG TGTTCATGC TTCTCGAGAT ACCCGATCA TATGAAACGG
    CGGTTGTGAA CAGTGATGAA AGACAATACC ACAAGTTACG AAGAGCTCTA TGGGTCTAGT ATACTTTGCC

GFP with introns
=====
491 CATGACTTTT TCAAGAGTGC CATGCCCGAA GGTATGTAC AGGAAAGAAC TATATTTTTC AAAGATGACG
    GTACTGAAAA AGTTCTCAGC GTACGGGCTT CCAATACATG TCCTTTCTTG ATATAAAAAAG TTTCTACTGC

GFP with introns
=====
561 GGAATACAA GACACGTAG TTTAAGCAGT TCGGTACTAA CTAACCATAC ATATTTAAAT TTTCAAGTGC
    CCTGTATGTT CTGTGCATTC AAATTTGTCA AGCCATGATT GATTGGTATG TATAAATTA AAAGTCCACG

GFP with introns
=====
631 TGAATCAAG TTTGAAGGTG ATACCCTTGT TAATAGAATC GAGTTAAAAG GTATTGATT TAAAGAAGAT
    ACTTCAGTTC AAACCTCCAC TATGGGAACA ATTATCTTAG CTCAATTTTC CATAACTAAA ATTTCTTCTA

GFP with introns
=====
701 GGAAACATTC TTGGACACAA ATTGGAATAC AACTATAACT CACACAATGT ATACATCATG GCAGACAAAC
    CCTTTGTAAG AACCTGTGTT TAACCTTAGT TTGATATTGA GTGTGTACA TATGTAGTAC CGTCTGTTG

GFP with introns
=====
771 AAAAAGATGG AATCAAGTT GTAAGTTTAA ACTTGGACTT ACTAACAAC GGATTATATT TAAATTTTCA
    TTTTCTTACC TTAGTTTCAA CATTCAAAT TGAACCTGAA TGATTGATTG CCTAATATAA ATTTAAAGT

GFP with introns
=====
841 GAACITCAAA ATTAGACACA ACATTGAAGA TGAAGCGTT CAACTAGCAG ACCATTATCA ACAAATACT
    CTTGAAGTTT TAATCTGTGT TGTAACCTCT ACCTTCGCAA GTTGATCGTC TGGTAATAGT TGTTTTATGA

GFP with introns
=====
911 CCAATTGCGG ATGGCCCTGT CTTTACCA GACAACCACT ACCTGTCCAC ACAATCTGCC CTTTCGAAAG
    GGTAAACCGC TACCGGACA GGAATGGT CTGTTGTAA TGGACAGGTG TGTTAGACGG GAAAGCTTT

GFP with introns
=====
981 ATCCCAACGA AAAGAGAGAC CACATGGTCC TTCTTGAGTT TGTAACAGCT GCTGGGATTA CACATGGCAT
    TAGGGTTGCT TTTCTCTCTG GTGTACCAGG AAGAACTCAA ACATTGTGCA CGACCTAAT GTGTACCCTA

```

## FIG. 6 (CONTINUED 1.)

GFP with introns  
unc-54 3' UTR

1051 GGATGAACTA TACAAATAGG GCGGCGCGAG CTCGCGATCG GCGGCTGTCA TCAGATCGCC ATCTCGCGCC  
CCTACTTGAT ATGTTTATCC CGGCGCGCTC GAGGCGTAGC CGGCGACAGT AGTCTAGCGG TAGAGCGCGG

unc-54 3' UTR

1121 CGTGCTCTG ACTTCTAAGT CCAATTACTC TTCAACATCC CTACATGCTC TTTCTCCCTG TGCTCCACCC  
GCACGGAGAC TGAAGATTCA GGTTAATGAG AAGTTGTAGG GATGTACGAG AAAGAGGGAC ACGAGGGTGG

unc-54 3' UTR

1191 CCCTATTTT GTTATTATCA AAAAACTTC TTCTTAATTT CTTTGTTTT TAGCTTCTT TAAGTCACCT  
GGGATAAAAA CAATAATAGT TTTTGTGAAG AAGAATTAAA GAAACAAAA ATCGAAGAAA ATTCAGTGGG

unc-54 3' UTR

1261 CTAACAATGA AATTGTGTAG ATTCAAAAT AGAATTAAT CGTAATAAAA AGTCGAAAAA AATTGTGCTC  
GATTGTACT TTAACACATC TAAGTTTTA TCCTAATTAA GCATTATTTT TCAGCTTTTT TTAACACGAG

unc-54 3' UTR

1331 CCTCCCCCA TTAATAATAA TTCTATCCCA AAATCTACAC AATGTTCTGT GTACACTTCT TAGTGTTTTT  
GGAGGGGGGT AATTATTATT AAGATAGGGT TTTAGATGTG TTACAAGACA CATGTGAAGA ATACAAAAA

unc-54 3' UTR

1401 TTACTTCTGA TAAATTTTTT TTGAACATC ATAGAAAAA CCGCACACAA AATACCTTAT CATATGTTAC  
AATGAAGACT ATTAAAAA AACTTTGTAG TATCTTTTTT GCGGTGTGTT TTATGGAATA GTATACAATC

unc-54 3' UTR

1471 GTTTCAGTTT ATGACCGCAA TTTTATTTT TTGCGACGTC TGGGCTCTC ATGACGTC AAATCATGCTCA  
CAAAGTCAAA TACTGGCGTT AAAAATAAAG AAGCGTGACG ACCCGGAGAG TACTGCAGT TAGTACGAGT

unc-54 3' UTR

1541 TCGTGAAAAA GTTTTGGAGT ATTTTGGAA TTTTCAATC AAGTGAAAGT TTATGAAAT AATTTTCCTG  
AGCACTTTTT CAAACCTCA TAAAACTT AAAAAGTTAG TTCACTTTCA AATACITTA TAAAAAGGAC

unc-54 3' UTR

1611 CTTTTCCTTT TTGGGGGTTT CCCCTATTGT TTGTCAAGAG TTTCGAGGAC GCGGTTTTTC TTGCTAAAT  
GAAACGAAA AACCCCAA GGGGATAACA AACAGTTCTC AAGCTCCTG CCGCAAAAG AACGATTTTA

unc-54 3' UTR

1681 CACAAGTATT GATGAGCAGC ATGCAAGAAA GATCGGAAGA AGGTTTGGGT TTGAGGCTCA GTGGAAGGTG  
GTGTCATAA CTACTCGTGC TACGTTCTTT CTAGCCTTCT TCCAACCCA AACTCCGAGT CACCTTCCAC

unc-54 3' UTR

1751 AGTAGAAGTT GATAATTTGA AAGTGGAGTA GTGTCTATGG GGTTTTGGC TTAAATGACA GAATACATTC  
TCATCTCAA CTATTAACT TTCACCTCAT CACAGATACC CAAAAACGG AATTTACTGT CTTATGTAAG

unc-54 3' UTR

1821 CCAATATACC AAACATAACT GTTTCCTACT AGTCGGCCGT ACGGGCCCTT TCGTCTCGCG CGTTTCGGTG  
GGTTATATGG TTGTATTGA CAAGGATGA TCAGCCGGCA TGCCCGGGA AGCAGAGCGC GCAAGGCCAC

1891 ATGACGGTGA AAACCTCTGA CACATGCAGC TCCCGGAGAC GGTACAGCT TGCTGTAAAG CGGATGCCGG  
TACTGCCACT TTTGGAGACT GTGTACGTGC AGGCGCTCTG CCAGTGTGCA ACAGACATTC GCCTACGGCC

1961 GAGCAGACAA GCCCGTCAGG GCGCGTCAGC GGGTGTGGC GGGTGTGGG GCTGGCTTAA CTATGCGGCA  
CTCGTCTGTT CGGCGAGTCC CGCGCAGTGC CCCACAACCG CCCACAGCCC CGACCGAAT GATACGCCGT

2031 TCAGAGCAGA TTGTAAGTGA AGTGACCATC ATGCGGTGTG AAATACCGCA CAGATGCGTA AGGAGAAAAT  
AGTCTCGTCT AACATGACTC TCACGTGGA TACGCCACAC TTTATGGCGT GTCTACGCAT TCCTCTTTTA

2101 ACCGCATCAG GCGGCCTTAA GGGCCTCGTG ATACGCCTAT TTTTATAGGT TAATGTCTAG ATAATAATGG  
TGGCGTAGTC CGCGGAAT CCCGGAGCAC TATGCGGATA AAAATATCCA ATTACAGTAC TATTATTACC

## FIG. 6 (CONTINUED 2.)

2171 TTCTTAGAC GTCAGGTGGC ACTTTTCGGG GAAATGTGCG CGGAACCCCT ATTTGTTTAT TTTCTAAAT  
AAAGAATCTG CAGTCCACCG TGAAAAGCCC CTTTACACGC GCCTTGGGGA TAAACAAATA AAAAGATTTA

2241 ACATTCAAAAT ATGTATCCGC TCATGAGACA ATAACCCTGA TAAATGCTTC AATAATATTG AAAAAGGAAG  
TGTAAGTTTA TACATAGGCG AGTACTCTGT TATTGGGACT ATTTACGAAG TTATTATAAC TTTTCCTTC

amp

2311 AGTATGAGTA TTCAACATTT CCGTGTGCGC CTTATTCCCT TTTTTCGGC ATTTTGCCTT CCTGTTTTTG  
TCATACTCAT AAGTTGTAAA GGCACAGCGG GAATAAGGGA AAAACGCGG TAAACGGAA GGACAAAAAC

amp

2381 CTCACCCAGA AACGCTGGTG AAAGTAAAG ATGCTGAAGA TCAGTTGGGT GCACGAGTGG GTTACATCGA  
GAGTGGGTCT TTGCGACCAC TTTCAATTTT TACGACTTCT AGTCAACCCA CGTGCTCACC CAATGTAGCT

amp

2451 ACTGATCTC AACAGCGGTA AGATCCTGA GAGTTTTCG CCGAAGAAG GTTTTCCAAT GATGAGCACT  
TGACCTAGAG TTGTCGCCAT TCTAGGAAC CTCAAAGCG GGGCTTCTTG CAAAGGTTA CTACTCGTGA

amp

2521 TTTAAGTTC TGCTATGTGG CGCGGTATTA TCCCGTATTG ACGCCGGGCA AGAGCACTC GGTGCGGCA  
AAATTTCAAG ACGATACACC GCGCCATAAT AGGGCATAAC TGCGGCCCGT TCTCGTTGAG CCAGCGGCGT

amp

2591 TACACTATTC TCAGAAAGAC TTGGTTGAGT ACTCACCAGT CACAGAAAAG CATCTTACGG ATGGCATGAC  
ATGTGATAAG AGTCTTACTG AACCAACTCA TGAGTGGTCA GTGTCTTTC GTAGAAAGCC TACCGTACTG

amp

2661 AGTAAGAGAA TTATGCAGTG CTGCCATAAC CATGAGTGAT AACACTGCGG CCAACTTACT TCTGACAACG  
TCATTCTCTT AATACGTCAC GACGGTATTG GTACTCACTA TTGTGACGCC GGTGAATGA AGACTGTTGC

amp

2731 ATCGGAGGAC CGAAGGAGCT AACCGCTTTT TTGCACAACA TGGGGGATCA TGTAACCTGC CTTGATCGTT  
TAGCCTCCTG GCTTCTCGA TTGGCGAAA AACGTGTTGT ACCCCCTAGT ACATTGAGCG GAAC TAGCAA

amp

2801 GGGAACCGGA GCTGAATGAA GCCATACCAA ACGACGAGCG TGACACCACG ATGCCTGTAG CAATGGCAAC  
CCCTTGGCCT CGACTTACTT CGGTATGTTT TGCTGCTCGC ACTGTGGTGC TACGGACATC GTTACCGTTG

amp

2871 AACGTTGCGC AACTATTAA CTGGCGAACT ACTTACTCTA GCTTCCCGGC AACAAATTAAT AGACTGGATG  
TTGCAACGCG TTGATAATT GACCGCTTGA TGAATGAGT CGAAGGGCCG TTGTTAATTA TCTGACCTAC

amp

2941 GAGGCGGATA AAGTTGCAGG ACCACTTCTG CGCTCGGCCC TTCCGGCTGG CTGGTTTATT GCTGATAAAT  
CTCGCCTAT TTCAACGTCC TGGTGAAGAC GCGAGCCGGG AAGGCCGACC GACCAATAA CGACTATTTA

amp

3011 CTGAGCGCGG TGAGCGTGGG TCTCGCGGTA TCATTGCAGC ACTGGGGCCA GATGGTAAGC CCTCCCGTAT  
GACCTCGGCC ACTCGCACC AGAGCGCCAT AGTAACGTG TGACCCCGGT CTACCATTCG GGAGGGCATA

amp

3081 CGTAGTTATC TACACGACGG GGAGTCAGGC AACTATGGAT GAACGAAATA GACAGATCGC TGAGATAGGT  
GCATCAATAG ATGTGCTGCC CTTAGTCCG TTGATACCTA CTTGCTTTAT CTGTCTAGCG ACTCTATCCA

## FIG. 6 (CONTINUED 3.)

amp  
-----

3151 GCCTCACTGA TTAAGCATTG GTAACGTCA GACCAAGTTT ACTCATATAT ACTTTAGATT GATTTAAAC  
CGGAGTGACT AATTCGTAAC CATTGACAGT CTGGTTCAAA TGAGTATATA TGAAATCTAA CTAATTTTG

3221 TTCATTTTTA ATTTAAAGG ATCTAGGTGA AGATCCTTTT TGATAATCTC ATGACCAAAA TCCCTTAACG  
AAGTAAAAAT TAAATTTCC TAGATCCACT TCTAGGAAAA ACTATTAGAG TACTGGTTTT AGGGAATTGC

3291 TGAGTTTTCG TTCCACTGAG CGTCAGACCC CGTAGAAAAG ATCAAAGGAT CTCTTGAGA TCCTTTTTTT  
ACTCAAAAGC AAGGTGACTC GCAGTCTGGG GCATCTTTTC TAGTTTCTTA GAAGAATCTT AGGAAAAAAA

3361 CTGCGCGTAA TCTGCTGCTT GCAAACAAA AAACCACCGC TACCAGCGGT GGTGTGTTG CCGGATCAAG  
GACGCGCATT AGACGACGAA CGTTTGTTC TTTGGTGGCG ATGGTCGCCA CCAAACAAAC GGCCTAGTTC

3431 AGCTACCAAC TCCTTTTCCG AAGGTAACGT GCTTCAGCAG AGCGCAGATA CCAAATACTG TCCTTCTAGT  
TCGATGGTTG AGAAAAGGC TTCCATTGAC CGAAGTCGTC TCGCGTCTAT GGTATTATGAC AGGAAGATCA

3501 GTAGCCGTAG TTAGGCCACC ACTTCAAGAA CTCTGTAGCA CCGCCTACAT ACCTCGCTCT GCTAATCCTG  
CATCGGCATC AATCCGGTGG TGAAGTTCTT GAGACATCGT GGCGGATGTA TGGAGCGAGA CGATTAGGAC

3571 TTACCAGTGG CTGCTGCCAG TGGCGATAAG TCGTGTCTTA CCGGGTTGGA CTCAAGACGA TAGTTACCGG  
AATGGTCACC GACGACGGTC ACCGCTATTG AGCACAGAAT GGCCCAACCT GAGTTCTGCT ATCAATGGCC

3641 ATAAGGCGCA GCGGTCGGCC TGAACGGGGG GTTCGTGCAC ACAGCCCAGC TTGGAGCGAA CGACCTACAC  
TATTCCCGCT CGCCAGCCCG ACTTGCCCC CAAGCACGTG TGTCCGGTCG AACCTCGCTT GCTGGATGTG

3711 CGAACTGAGA TACCTACAGC GTGAGCATTG AGAAAGCGCC ACGCTTCCCG AAGGGAGAAA GCGCGACAGG  
GCTTGACTCT ATGGATGTCG CACTCGTAAC TCTTTCGCGG TGCGAAGGGC TTCCCTCTTT CCGCCTGTCC

3781 TATCCGGTAA GCGGCAGGGT CGGAACAGGA GAGCGCACGA GGGAGCTTCC AGGGGGAAC SCCTGSTATC  
ATAGGCCATT CGCGTCCCA GCCTTGTCCT CTCGCGTCTC CCCTCGAAGG TCCCCCTTG CGGACCATAG

3851 TTTATAGTCC TGTCGGGTTT CGCCACCTCT GACTTGAGCG TCGATTTTG TGATGCTCGT CAGGGGGGCG  
AAATATCAGG ACAGCCCAA GCGGTGGAGA CTGAACCTCG AGCTAAAAAC ACTACGAGCA GTCCCCCGC

3921 GAGCCTATGG AAAACGCCA GCAACGCGGC CTTTTACGG TTCCTGGCCT TTTGCTGGCC TTTTGCTCAC  
CTCGGATACC TTTTGCGGT CGTTGCGCG GAAAAATGCC AAGGACCGGA AAACGACCGG AAAACGAGTG

3991 ATGTTCTTTC CTGCGTTATC CCCTGATTCT GTGGATAACC GTATTACCGC CTTTGAGTGA GCTGATACCG  
TACAAGAAAG GACGCAATAG GGGACTAAGA CACCTATTGG CATAATGGCG GAAACTCACT CGACTATGGC

4061 CTCGCCGAG CCGAACGACC GAGCGCAGCG AGTCAGTGAG CGAGGAAGCG GAAGAGCGCC CAATACGCAA  
GAGCGGCGTC GGCTTGCTGG CTCGCGTCGC TCAGTCATC GCTCCTTCGC CTTCTCGCGG GTTATGCGTT

4131 ACCGCCTCTC CCCGCGGTT GCGCGATTCA TTAATGCAGC TGGCACGACA GGTTCCCGA CTGGAAAGCG  
TGGCGGAGAG GGGCGCGCAA CCGGCTAAGT AATTACGTG ACCGTGCTGT CCAAAGGGCT GACCTTTCGC

4201 GGCAGTGAGC GCAACGCAAT TAATGTGAGT TAGCTCACTC ATTAGGCACC CCAGGCTTTA CACTTTATGC  
CCGTCACTCG CGTTGCGTTA ATTACACTCA ATCGAGTGAG TAATCCGTGG GGTCCGAAAT GTGAAATACG

4271 TTCCGGCTCG TATGTTGTGT GGAATTGTGA GCGGATAACA ATTTACACA GGAAACAGCT ATGACCATGA  
AAGGCCGAGC ATACAACACA CCTTAACACT CGCCTATTGT TAAAGTGTGT CCTTGTGCGA TACTGGTACT

4341 TTACGCCA  
AATGCGGT

## SEQUENCE LISTING

&lt;110&gt; DEVGEN NV

&lt;120&gt; GENE EXPRESSION SYSTEM

&lt;130&gt; SCB/55177/001

&lt;140&gt;

&lt;141&gt;

&lt;160&gt; 5

&lt;170&gt; PatentIn Ver. 2.0

&lt;210&gt; 1

&lt;211&gt; 47

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

<223> Description of Artificial Sequence:  
oligonucleotide o-GN59

&lt;400&gt; 1

ctagattaca actaattata cttatttgaa tattcaaatt ttcagac

47

&lt;210&gt; 2

&lt;211&gt; 47

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

<223> Description of Artificial Sequence:  
oligonucleotide o-GN60

&lt;400&gt; 2

ccgggtctga aaatttgaat attcaaataa gtataattag ttgtaat

47

&lt;210&gt; 3

&lt;211&gt; 3498

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

<223> Description of Artificial Sequence: plasmid  
pDW3123

&lt;400&gt; 3

```

agcttggcgc ctaatacgac tcactatagg gctgcaggtc gactctagat tacaactaat 60
tatacttatt tgaatattca aattttcaga cccgggattg gccaaaggac ccaaaggat 120
gtttcgaatg atactaacat aacatagaac attttcagga ggacccttgg ctacgctcct 180
gctgggatta cacatggcat ggatgaacta tacaatatagg gccggccgag ctccgcatcg 240
gccgctgtca tcagatcgcc atctcgcgcc cgtgcctctg acttctaagt ccaattactc 300
ttcaacatcc ctacatgctc tttctccctg tgctcccacc ccctattttt gttattatca 360
aaaaaacttc ttcttaattt ctttgttttt tagcttcttt taagtcacct ctaacaatga 420
aattgtgtag attcaaaaat agaattaatt cgtaataaaa agtcgaaaaa aattgtgctc 480
cctcccccca ttaataataa ttctatccca aaatctacac aatgttctgt gtacacttct 540
tatgtttttt ttacttctga taaatttttt ttgaaacatc atagaaaaaa ccgcacacaa 600
aataccttat catatgttac gtttcagttt atgaccgcaa tttttatttc ttgcacgctc 660
tgggcctctc atgacgtcaa atcatgctca tcgtgaaaaa gttttggagt atttttggaa 720

```

```

tttttcaatc aagtgaaggt ttatgaaatt aattttcctg cttttgcttt ttggggggttt 780
cccctattgt ttgtcaagag ttctgaggac ggcgtttttc ttgctaaaaat cacaagtatt 840
gatgagcacg atgcaagaaa gatcggaaga aggtttgggt ttgaggctca gtggaagggtg 900
agtagaaggt gataatttga aagtggagta gtgtctatgg ggtttttgcc ttaaatgaca 960
gaatacattc ccaatatacc aaacataact gtttcctact agtcggccgt acgggccctt 1020
tcgtctcgcg. cgtttcggtg atgacggtga aaacctctga cacatgcagc tcccgagac 1080
ggtcacagct tgtctgtaag cggatgccgg gagcagacaa gcccgtcagg gcgcgtcagc 1140
gggtgttggtc ggggtgcggg gctggcttaa ctatgcggca tcagagcaga ttgtactgag 1200
agtgacccat atgcgggtgtg aaataccgca cagatgcgta aggagaaaat accgcatcag 1260
gcggccttaa gggcctcgtg atacgcctat ttttataggt taatgtcatg ataataatgg 1320
tttcttagac gtacgggtggc acttttcggg gaaatgtgcg cggaaacctt atttgtttat 1380
ttttctaaat aatacaaat atgtatccgc tcatgagaca ataacctga taaatgcttc 1440
aataatattg aaaaagggaag agtatgagta ttcaacattt ccgtgtcgcc cttattccct 1500
tttttgcggc attttgcctt cctgtttttg ctcaccacaga aacgctggtg aaagtaaaag 1560
atgctgaaga tcagttgggt gcacgagtggt gttacatcga actggatctc aacagcggta 1620
agatccttga gagttttcgc cccgaagaac gttttccaat gatgagcact tttaaagttc 1680
tgctatgtgg cgcggtatta tcccgatttg acgcccggca agagcaactc ggtcgccgca 1740
tacactatc tcagaatgac ttggttgagt actcaccagt cacagaaaag catcttacgg 1800
atggcatgac agtaagagaa ttatgcagtg ctgccataac catgagtgat aacactgcgg 1860
ccaacttact tctgacaacg atcggaggac cgaaggagct aaccgctttt ttgcacaaca 1920
tgggggatca tgtaactcgc cttgatcgtt gggaaccgga gctgaatgaa gccataccaa 1980
acgacgagcg tgacaccacg atgcctgtag caatggcaac aacgttgcg c aaactattaa 2040
ctggcgaaact acttactcta gcttcccggc aacaattaat agactggatg gaggcggata 2100
aagttgcagg accacttctg cgtcggccc ttccggcttg ctggtttatt gctgataaat 2160
ctggagccgg tgagcgtggg tctcgcggtg tcattgcagc actggggcca gatggtaagc 2220
cctcccgtat cgtagttatc tacacgacgg ggagtcagcg aactatggat gaacgaaata 2280
gacagatcgc tgagataggt gcctcactga ttaagcattg gtaactgtca gaccaagttt 2340
actcatatat acttttagatt gatttaaaac ttcatTTTTa atttaaaagg atctaggtga 2400
agatcctttt tgataatctc atgacaaaaa tcccttaacg tgagttttcg ttccactgag 2460
cgtcagacc cgtagaaaaag atcaaaaggat cttcttgaga tccttttttt ctgcgcgtaa 2520
tctgctgctt gcaaacaaaa aaaccaccgc taccagcggg ggtttgtttg ccggatcaag 2580
agctaccaac tcttttttcg aaggtaactg gcttcagcag agcgcagata ccaataactg 2640
tccttctagt gtacccgtag ttaggccacc acttcaagaa ctctgtagca ccgcctacat 2700
acctcgctct gctaactctg ttaccagtgg ctgctgccag tggcgataag tcgtgtctta 2760
ccgggttgga ctcaagacga tagttaccgg ataaggcgca gcggtcgggc tgaacggggg 2820
gttcgtgcac acagcccagc ttggagcgaa cgacctacac cgaactgaga tacctacagc 2880
gtgagcattg agaaagcgcc acgcttcccg aaggagagaa ggcggacagg tatccggtaa 2940
gcggcagggt cggaacagga gagcgcacga gggagcttcc agggggaac gcctggtatc 3000
tttatagtc tgcggggtt cgccacctct gacttgagcg tcgatttttg tgatgctcgt 3060
cagggggggc gagcctatgg aaaaacgcca gcaacgcggc ctttttacgg ttcttgccct 3120
tttgcgtggc ttttgcac atgttcttct ctgcgttatc ccctgattct gtggataacc 3180
gtattaccgc ctttgagtga gctgataccg ctgcgcgag ccgaacgacc gagcgacgag 3240
agtcagtgag cgaggaagcg gaagagcgcc caatacgcaa accgcctctc ccgcgcggtt 3300
ggccgattca ttaatgcagc tggcacgaca ggtttcccga ctggaaagcg ggcagtgagc 3360
gcaacgcaat taatgtgagt tagctcactc attaggcacc ccaggcttta cactttatgc 3420
ttccggctcg tatgttgtgt ggaattgtga gcggataaca atttcacaca ggaacagct 3480
atgaccatga ttacgcca 3498

```

&lt;210&gt; 4

&lt;211&gt; 4348

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

<223> Description of Artificial Sequence: plasmid  
pDW3124

&lt;400&gt; 4

```

agcttggcgc ctaatacgac tcaactatagg gctgcaggtc gactctagat tacaactaat 60
tatacttatt tgaatatca aattttcaga cccgggattg gccaaaggac ccaaagggtat 120
gtttcgaatg atactaacat aacatagaac attttcagga ggaccttgg ctagcgtcga 180

```

cggtaccatg gggcgcgcca tgagtaaagg agaagaactt ttcactggag ttgtcccaat 240  
 tcttggtgaa ttagatgggtg atgttaatgg gcacaaatth tctgtcagtg gagagggtga 300  
 aggtgatgca acatacggaa aacttaccct taaatttatt tgcactactg gaaaactacc 360  
 tgttccatgg gtaagtttaa acatatatat actaactaac cctgattatt taaattttca 420  
 gccaacactt gtcactactt tctgttatgg tgttcaatgc ttctcgagat acccagatca 480  
 tatgaaacgg catgactttt tcaagagtgc catgcccga ggttatgtac aggaagaac 540  
 tatatttttc aaagatgacg ggaactacaa gacacgtaag tttaaacagt tccgtactaa 600  
 ctaaccatac atattttaa tttcaggtgc tgaagtcaag ttgaagggtg atacccttgt 660  
 taatagaatc gagttaaaag gtattgattt taaagaagat ggaaacattc ttggacacaa 720  
 attggaatac aactataact cacacaatgt atacatcatg gcagacaaac aaaagaatgg 780  
 aatcaaaagt gtaagttaa acttggactt actaactaac ggattatatt taaattttca 840  
 gaacttcaaa attagacaca acattgaaga tggaagcgtt caactagcag accattatca 900  
 acaaaatact ccaattggcg atggccctgt ccttttacc gacaaccatt acctgtccac 960  
 acaatctgcc ctttcgaaag atcccaacga aaagagagac cacatgggtcc ttcttgagtt 1020  
 tgtaacagct gctgggatta cacatggcat ggatgaacta tacaatagg gccggccgag 1080  
 ctccgcatcg gccgctgtca tcagatcgcc atctcgccgc cgtgcctctg acttctaagt 1140  
 ccaattactc ttcaacatcc ctacatgctc tttctccctg tgcctccacc ccctattttt 1200  
 gttattatca aaaaaacttc ttcttaattt ctttgttttt tagcttcttt taagtccact 1260  
 ctaacaatga aattgtgtag attcaaaat agaattaatt cgtaataaaa agtcgaaaaa 1320  
 aatttggctc cttcccccac ttaataataa ttctatccca aaatctacac aatgttctgt 1380  
 gtacacttct tatgtttttt ttacttctga taaatttttt ttgaacatc atagaaaaaa 1440  
 ccgcacacaa aataccttat catatgttac gtttcagttt atgaccgcaa tttttatttc 1500  
 ttgcgacgtc tgggcctctc atgacgtcaa atcatgctca tccgtaaaaa gttttggagt 1560  
 atttttggaa tttttcaatc aagtgaagtt ttatgaattt aattttcctg cttttgcttt 1620  
 ttgggggttt cccctattgt ttgtcaagag ttctcaggac ggcgtttttc ttgctaaaat 1680  
 cacaagtatt gatgagcacg atgcaagaaa gatcggaaga aggtttgggt ttgaggctca 1740  
 gtggaagggt agtagaagtt gataatttga aagtggagta gtgtctatgg ggtttttgcc 1800  
 ttaaagtaca gaatacattc ccaatatacc aaacataact gtttcctact agtcggccgt 1860  
 acgggccctt tccgtctcgc cgtttcgggtg atgacggtga aaacctctga cacatgcagc 1920  
 tcccgagac ggtcacagct tgtctgtaag cggatgccgg gagcagacaa gcccgtcagg 1980  
 gcgcgtcagc ggggtgttggc ggggtgtcggg gctggcctaa ctatcgcgca tcagagcaga 2040  
 ttgtactgag agtgcacatc atgcccgtgt aaataccgca cagatgcgta aggagaaaat 2100  
 accgcatcag gcggccttaa gggcctcgtg atacgcctat ttttatagtt taatgtcatg 2160  
 ataataatgg tttcttagac gtcagggtggc acttttcggg gaaatgtgcg cggaaaccct 2220  
 atttgtttat ttttctaaat acattcaaat atgtatccgc tcatgagaca ataaccctga 2280  
 taaatgcttc aataatattg aaaaaggaag agtatgagta ttcaacattt cctgttcgcc 2340  
 cttattccct tttttcgccg cctgtttttg ctcaccacga aacgctggtg 2400  
 aaagtaaaag atgctgaaga tcagttgggt gcacgagtg gttacatcga actggatctc 2460  
 aacagcggta agatccttga gagttttcgc cccgaagaac gttttccaat gatgagcact 2520  
 tttaaagttc tgctatgtgg cgcggtatta tcccgtattg acgcccggca agagcaactc 2580  
 ggtcgccgca tcacactatc tcagaatgac ttggttgagt actcaccagt cacagaaaag 2640  
 catcttaccg atggcatgac agtaagagaa ttatgcagtg ctgccataac catgagtgat 2700  
 aacactgcgg ccaacttact tctgacaacg atcggaggac cgaaggagct aaccgctttt 2760  
 ttgcacaaca tgggggatca tgtaactcgc cttgatcgtt gggaaccgga gctgaatgaa 2820  
 gccataccaa acgacgagcg tgacaccacg atgcctgtag caatggcaac aacgttgcgc 2880  
 aaactattaa ctggcgaact acttactcta gcttcccggc aacaattaat agactggatg 2940  
 gagggcgata aagttgcagg accacttctg cgctcggccc ttccggctgg ctggtttatt 3000  
 gctgataaat ctggagccgg tgagcgtggg tctcgcggta tcatgtcagc actggggcca 3060  
 gatgtaagc cctcccgatc cgtagtattc tacacgacgg ggagtcaggc aactatggat 3120  
 gaacgaaata gacagatcgc tgagataggt gcctcactga ttaagcattg gtaactgtca 3180  
 gaccaagttt actcatatat actttagatt gatttaaaac ttcatthttt atttaaaagg 3240  
 atctaggtga agatcctttt tgataatctc atgacaaaaa tcccttaacg tgagttttcg 3300  
 ttccactgag cgtcagaccc cgtagaaaaa atcaaaggat cttcttgaga tccctttttt 3360  
 ctgcgcgtaa tctgtgctt gcaaacacac aaaccaccgc taccagcggg ggtttgtttg 3420  
 ccggatcaag agctaccaac tctttttccg aaggttaact gcttcagcag agcgagata 3480  
 ccaaaactag tcttcttagt gtagccgtag ttaggccacc acttcaagaa ctctgtagca 3540  
 ccgcctacat acctgcctct gctaactcct ttaccagtgg ctgctgccag tggcgataag 3600  
 tccgtgtctta ccgggttgga ctcaagacga tagttaccgg ataaggcgca gcggtcgggc 3660  
 tgaacggggg gttcgtgcac acagcccagc ttggagcgaa cgacctacac cgaactgaga 3720  
 tactacagc tttagcattg agaaagcgcc acgcttcccg aaggagaaaa ggcggacagg 3780  
 tatccggtaa gcggcagggt cggaaacagga gagcgacaga gggagcttcc agggggaaac 3840

```

gcctgggtatc tttatagtc tgtcgggttt cgccacctct gacttgagcg tcgatttttg 3900
tgatgctcgt caggggggag gagcctatgg aaaaacgcca gcaacgcggc ctttttacgg 3960
ttcctggcct tttgctggcc ttttgctcac atgttctttc ctgcttate ccctgattct 4020
gtggataacc gtattaccgc ctttgagtga gctgataccg ctgcgcgag ccgaacgacc 4080
gagcgacgag agtcagttag cgaggaagcg gaagagcgcc caatacgcaa accgcctctc 4140
cccgcgcggtt ggccgattca ttaatgcagc tggcacgaca ggtttccgca ctggaagcg 4200
ggcagtgage gcaacgcaat taatgtgagt tagctcactc attaggcacc ccaggcttta 4260
cactttatgc ttccggctcg tatgttgtgt ggaattgtga gcggataaca atttcacaca 4320
ggaacagct atgaccatga ttacgcca 4348

```

&lt;210&gt; 5

&lt;211&gt; 9309

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

&lt;223&gt; Description of Artificial Sequence: plasmid pGN148

&lt;400&gt; 5

```

atgactgctc caaagaagaa gcgtaaggta ccggtaatga acacgattaa catcgctaag 60
aacgacttct ctgacatcga actggctgct atcccgttca acactctggc tgaccattac 120
ggtgagcggt tagctcggtg agtttaaaaca tctagatact aactaacgat taacatttaa 180
attttcagcg aacagttggc ccttgagcat gactcttacg agatgggtga agcacgcttc 240
cgcaagatgt ttgagcgta acttaagct ggtgaggttg cggataacgc tgccgccaag 300
cctctcatca ctaccctact ccctaagatg attgcacgca tcaacgactg gtttgaggaa 360
gtgaaagcta agcgcggcaa gcgcccagca gccttccagt tcctgcaaga aatcaagccg 420
gaagccgtag cgtacatcac cattaagacc actctggctt gcctaaccag tgcagacaat 480
acaaccgttc aggtcttagc aagcgcaatc ggtcgggcca ttgaggacga ggctcgcttc 540
ggtcgtatcc gtgaccttga agctaagcac ttcaagaaaa acgttgagga acaactcaac 600
aagcgcgtag ggcacgtcta caagaaagca tttatgcaag ttgtcgaggc tgacatgctc 660
tctaagggtc tactcggtgg cgagcggtgg tcttcgtggc ataaggaaga ctctattcat 720
gtaggagtac gtgcatcga gatgctcatt gactcaaccg gactgggttag cttacaccgc 780
caaaatgctg gcgtagtagg tcaagactct gagactatcg aactcgacc tgaatacgct 840
gaggctatcg caaccgtgc aggtgcgctg gctggcatct ctccgatgtt ccaaccttgc 900
gtagttcctc ctaagccgtg gactggcatt actggtgggt gctattgggc taacggctcg 960
cgtcctctgg cggtggtg tctcacagt aagaaagcac tgatgcgcta cgaagacgtt 1020
tacatgcctg aggtgtacaa agcgattaac attgcgcaa acacgcgatg gaaaatcaac 1080
aagaaagctc tagcgtgcgc caacgtaatc accaagtggg agcattgtcc ggtcgaggac 1140
atccctgcga ttgagcgtag agaactcccg atgaaaccgg aagacatcga catgaatcct 1200
gaggctctca ccgctggaa acgtgctgcc gctgctgtgt accgcaagga caaggctcgc 1260
aagttctgcc gtatcagcct tgagttcatg cttgagcaag ccaataagtt tgctaaccat 1320
aaggccatct ggttccctta caacatggac tggcgcggtc gtgtttacgc tgtgtcaatg 1380
ttcaaccgcg aagctaacga tatgaccaa ggactgctta cgctggcgaa aggtaaacca 1440
atcggtaaag aaggttacta ctggctgaaa atccacggtg caaactgtgc ggggtgtcgat 1500
aaggttccgt tccctgagcg catcaagttc attgaggaaa accacgagaa catcatggct 1560
tgcgctaagt ctccactgga gaacacttgg tgggctgagc aagattctcc gttctgcttc 1620
cttgcttctt gctttgagta cgctggggtg cagcaccacg gcctgagcta taactgctcc 1680
cttccgctgg cgtttgacgg gtcttgctct ggcatccagc actctccgc gatgctccga 1740
gatgaggtag gtgctgcgc gggtgtaagt ttaactcta tctactaac taacgaagct 1800
tatttaaatt ttcagaactt gcttcctagt gaaaccgttc aggacatcta cgggattgtt 1860
gctaagaaag tcaacgagat tctacaagca gacgcaatca atgggaccga taacgaagta 1920
gttaccgtga ccgatgagaa cactggtgaa atctctgaga aagtcaagct gggcactaag 1980
gcactggctg gtcaatggct ggcttacggg gttactcgca gtgtgactaa gcgttcagtc 2040
atgacgctgg cttacgggtc caaagagttc ggcttccgtc aacaagtgtt ggaagatacc 2100
attcagccag ctattgattc cggcaagggt ctgatgttca ctacgccaag tcaggctgct 2160
ggatacatgg ctaagctgat ttgggaatct gtgagcgtag cgggtgtagc tgcggttgaa 2220
gcaatgaact ggcttaagtc tgctgctaag ctgctggctg ctgaggtcaa agataagaag 2280
actggagaga ttctcgcaa gcgttgcgct gtgcattggg tcaactccga tggtttccct 2340
gtgtggcagg aatacaagaa gcctattcaa acgcgtttga acctgatgtt cctcggtcag 2400
ttccgcttac agcctacat taacaccaac aaagatagcg agattgatgc acacaaacag 2460
gagtctggtg tcgctcctaa cttgttacac agccaagacg gtagccacct tcgtaagact 2520

```

gtagtggtggg cacacgagaa gtacggaatc gaatcttttg cactgattca cgactccttc 2580  
 ggtaccattc cggctgacgc tgcgaacctg ttcaaagcag tgcgcgaaac tatggttgac 2640  
 acatatgagt cttgtgatgt actggctgat ttctacgacc agttcgctga ccagttgcac 2700  
 gtagtctcaat tggacaaaat gccagcactt ccggctaaag gtaacttgaa cctccgtgac 2760  
 atcttagagt cggacttcgc gttcgcgtaa gaattccaac tgagcgccgg tcgctaccat 2820  
 taccacttg tctggtgtca aaaataatag gggccgctgt catcagagta agtttaaaact 2880  
 gagtctact aactaacgag taatatatta attttcagca tctcgcgccc gtgcctctga 2940  
 cttctaagtc caattactct tcaacatccc tacatgctct ttctccctgt gctcccacc 3000  
 cctatttttg ttattatcaa aaaaacttct tcttaatttc tttgttttt agcttctttt 3060  
 aagtcacctc taacaatgaa attgtgtaga ttcaaaaata gaattaattc gtaataaaaa 3120  
 gtcgaaaaaa attgtgctcc ctcccccat taataataat tctatcccaa aatctacaca 3180  
 atgttctgtg tacacttctt atgttttttt tacttctgat aaattttttt tgaacatca 3240  
 tagaaaaaac cgcacacaaa ataccttatt atatgttacg tttcagttta tgaccgcaat 3300  
 ttttatttct tcgcacgtct gggcctctca tgacgtcaaa tcatgctcat cgtgaaaaag 3360  
 ttttgagta tttttggaat ttttcaatca agtgaaagt tatgaaatta attttcctgc 3420  
 tttgtctgt tgggggttt ccctattgtt tgcaagagt ttcgaggacg gcgtttttct 3480  
 tgctaaaatc accagattg atgagcacga tgcaagaaag atcggaagaa ggtttgggtt 3540  
 tgaggctcag tggaggtga gtagaagttg ataattgaa agtggagtag tgtctatggg 3600  
 gtttttgct taaatgacag aatacattcc caataacca aacataactg tttctacta 3660  
 gtcggccgta cgggcccttt cgtctcgcgc gtttcggtga tgacggtgaa aacctctgac 3720  
 acatgcagct ccgggagacg gtcacagcct gtctgtaagc ggatgccggg agcagacaag 3780  
 cccgtcaggg cgcgtcagcg ggtgtggcg ggtgtcgggg ctggctaac tatggggcat 3840  
 cagagcagat gtactgaga gtgcaccata tgcggtgtga aataccgcac agatgcgtaa 3900  
 ggagaaaata ccgcacatcag cggccttaag ggcctcgtga tacgcctatt tttataggtt 3960  
 aatgtcatga taataatggt ttcttagacg tcaggtggca cttttcgggg aaatgtgcgc 4020  
 ggaaccctta tttgtttatt tttctaaata cattcaaata tgtatccgct catgagacaa 4080  
 taaccctgat aaatgcttca ataatttga aaaaggaaga gtatgagat tcaacatttc 4140  
 cgtgtgcgcc ttattccctt ttttgcggca ttttgccttc ctgtttttgc tccccagaa 4200  
 acgctggtga aagtaaaaaga tgctgaagat cagttgggtg cacgagtggg ttacatcgaa 4260  
 ctggatctca acacggttaa gatccttgag agttttcggc ccgaagaacg ttttccaatg 4320  
 atgagcactt ttaaagttct gctatgtggc gcggtattat cccgtattga cgcggggcaa 4380  
 gagcaactcg gtgcgcgat acactattct cagaatgact tggttgagta ctacacagtc 4440  
 acagaaaagc atcttacgga tggcatgaca gtaagagaat tatgcagtgc tgccataacc 4500  
 atgagtata acacggttaa caacttact ctgacaacga tcggaggacc gaaggagcta 4560  
 accgcttttt tgcacaacat gggggatcat gtaactcgcc ttgatcgttg ggaaccggag 4620  
 ctgaatgaag ccataccaaa cgacgagcgt gacaccacga tgcctgtagc aatggcaaca 4680  
 acgttgccga aactattaac tggcgaacta cttactctag cttcccggca acaattaata 4740  
 gactggatgg agcggtataa agttgcagga ccacttctgc gctcggccct tcgggctggc 4800  
 tggtttattg ctgataaatc tggagccggt gagcgtgggt ctcgcggtat cattgcagca 4860  
 ctggggccag atggtaagcc ctcccgatc gtagttatct acacgacggg gagtacggca 4920  
 actatggatg aacgaaatag acagatcgct gagataggtg cctcactgat taagcattgg 4980  
 taactgtcag accaagttta ctcatatata ctttagattg atttaaaact tcatttttaa 5040  
 tttaaaagga tctaggtgaa gatccttttt gataatctca tgacaaaat cccttaacgt 5100  
 gagttttctg tccactgagc gtcagacccc gtagaaaaga tcaaaggatc ttcttgagat 5160  
 cctttttttc tgcgcgtaat ctgctgcttg caaacaaaa aaccaccgct accagcgtg 5220  
 gtttgtttgc cggatcaaga gctaccaact ctttttcgga aggttaactg cttcagcaga 5280  
 gcgcagatac caaataactgt ccttctagt tagccgtagt taggccacca cttcaagaac 5340  
 tctgtagcac cgctacata cctcgtctcg ctaatcctgt taccagtggc tgcgtccagt 5400  
 ggcgataagt cgtgtcttac cgggttgac tcaagacgat agttaccgga taaggcgag 5460  
 cggtcgggct gaacggggg ttcgtgcaca cagcccagct tggagcgaac gacctacacc 5520  
 gaactgagat acctacagcg tagcattga gaaagcgcca cgcttcccga agggagaaag 5580  
 gcggacaggt atccggtaag cggcagggtc ggaacaggag agcgcacgag ggagcttcca 5640  
 gggggaaaacg cctggtatct ttatagtcct gtcgggttc gccacctctg acttgagcgt 5700  
 cgatttttgt gatgctcgtc aggggggagg agcctatgga aaaacgccag caacgcggcc 5760  
 tttttacgt tcttggcctt ttgctggcct tttgtcaca tgttcttcc tgcgttatcc 5820  
 cctgattctg tccgtaaccg tattaccgcc tttgagttag ctgataccgc tcgccgcagc 5880  
 cgaacgaccg agcgcagcga gtcagtgagc gaggaagcgg aagagcgccc aatacgcaaa 5940  
 ccgcctctcc ccgcgcgttg gccgattcat taatgcagct ggcacgacag gtttcccagc 6000  
 tggaaagcgg gcagtgcgc caacgcaatt aatgtgagtt agctcactca ttaggcaccc 6060  
 caggctttac actttatgct tccggctcgt atgttgtgtg gaattgtgag cggataacaa 6120  
 tttcacacag gaaacagcta tgaccatgat tacgccaagc tgtaagtta aacatgatct 6180

tactaactaa ctattctcat ttaaattttc agagcttaaa aatggctgaa atcactcaca 6240  
acgatggata cgctaacaac ttggaaatga aataagcttg catgcctgca gagcaaaaaa 6300  
atactgcttt tctttgcaaa attcgggtgct ttcttcaaag agaaactttt gaagtcggcg 6360  
cgagcatttc ctcttttgac ttctctcttt ccgccaaaaa gcctagcatt tttattgata 6420  
atttgattac acacactcag agttcttcca catgataaag tgtttcattg gcactcgccc 6480  
taacagtaca tgacaaggcg ggattattat cgatcgatat tgaagacaaa ctccaaatgt 6540  
gtgctcattt tggagccccg tgtggggcag ctgctctcaa tatattacta gggagacgag 6600  
gagggggacc ttatcgaaag tgcgatgagc cattctttct tctttatgca ctctcttcac 6660  
tctctcacac attaatcgat tcatagactc ccatattcct tgatgaaggt gtgggttttt 6720  
agcttttttt ccgatttgtt aaaaggaaga ggctgacgat gttaggaaaa agagaacgga 6780  
gccgaaaaaa catccgtagt aagtcttctt ttttaagccga cacttttttag acagcattcg 6840  
ccgctagttt tgaagtttaa attttaaaaa ataaaaatta gtttcaattt tttttaatta 6900  
ctaaatagge aaaagttttt tcaagaactc tagaaaaact agcttaattc atgggtacta 6960  
gaaaaattct tgttttaaat ttaatattta tcttaagatg taattacgag aagctttttt 7020  
gaaaattctc aattaaaaa atttgccgat ttagaataaa agtcttcaga aatgagtaaa 7080  
agctcaaatc agaagtttgt ttttaaggga aaaacacgaa aaaagaacac tatttatctt 7140  
ttctctcccg cgtaaaaatta gttgtttgat taatagtgat ccgctgtcta tttgcactcg 7200  
gctcttcaca ccgtgcttcc tctcacttga cccaacagga aaaaaaaca tcacgtctga 7260  
gacgggtgaa tgcccttatca agagcgctcg ctctttcacc cagtaacaaa aaaaatttgg 7320  
tttctttact ttatatttat gtaggtcaca aaaaaaagt gatgcagttt tgtgggtcgg 7380  
ttgtctccac accacctccg cctccagcag cacacaatca tcttcgtgtg ttctcgacga 7440  
ttcttgtat gcccggtcg tgaatgcacc acattcgacg cgcaactaca caccacactc 7500  
actttcggtg gtattactac acgtcatcgt tgttcgtagt ctcccgtctt ttctgcccc 7560  
ctcactctc attattcccc ttggtgtatt gatttttttt aaatgggtaca ccactcctga 7620  
cgtttctacc tcttgttttt ccgtccattt agattttatc tggaaatttt tttaaaattt 7680  
taggccagag agttctagtt cttgttctaa aagtctaggt cagacataca tttctattt 7740  
ctcatcaaaa aaaaagtga taaagaaac tggttattca gaaagagtgt gtctcgttga 7800  
aattgattca aaaaaaatt cccaccctc gcttgtttct caaaatatga gatcaacgga 7860  
tttttctctt ctcgattcaa tttttgtcg cgtctgtctt gccaaagtgt gtgtgtccga 7920  
gcaaagatg agagaattta caaacagaaa tgaaaaaaag ttggccaaat aatgaagttt 7980  
tatccgagat tgatgggaaa gatattaatg ttctttacgg ttggagggg agagagagat 8040  
agattttcgc atcaaaactcc gccttttaca tgtcttttag aatctaaaat agatttttct 8100  
catcattttt aatagaaaaa cgagaaatta cagtaatttc gcaattttct tgccaaaaat 8160  
acacgaaatt tgtgggtctc gccacgatct cgggtcttagt ggttcatttg gtttaaaagt 8220  
ttataaaatt tcaaatctca gtgtttaatt tccgcataat tggacctaaa atgggttttt 8280  
gtcatcattt tcaacaagaa atcgtgaaaa tctctgtgtt tdcgaatttt cttttcaaaa 8340  
atacacgaaa tatatggtta tttcccgaat tattgaggtt ctccgccaga tttcagtcac 8400  
agtggccagg atttatcacg aaaaaagttc gcctagtctc acatttccgg aaaaccgaat 8460  
ctaaattagt tttttgtcat cattttgaac aaaaaatcga gacatcccta tagtttcgca 8520  
attttcgtcg cttttctctc caaaaatgac agtctagaat taaaattcgc tggaaactggg 8580  
accatgatat cttttctccc cgtttttcat tttatttttt attacactgg attgactaaa 8640  
ggtcaccacc accgccagtg tgtgccatat cacacacaca cacacacaca atgtcgagat 8700  
tttatgtgtt atccctgctt gatctcgttc cgttgtctct ctctctctat tcatctttt 8760  
agccgagaag ctccagagaa tggagcacac aggatcccgg cgcgcgatgt cgtcgggaga 8820  
tggcgccggc tgggaagccg ccgagagata tcaggaaga tegtctgatt tctcctcgga 8880  
tgccacctca tctctcgagt ttctccgctt gttactccct gccgaacctg atatttccc 8940  
ttgtcgtaaa gagatgtttt tattttactt tacaccgggt cctctctctc tgccagcaca 9000  
gctcagtggt ggctgtgtgc tggggtcctt gccaccggcg gcctcatctt cttcttcttc 9060  
ttctctctcg ctctcgctta tcaactcttc attcattctt attccttttc atcatcaaac 9120  
tagcatttct tactttattt atttttttca attttcaatt ttcagataaa accaaactac 9180  
ttgggttaca gccgtcaaca gatccccggg attggccaaa ggaccctaaag gtatgtttcg 9240  
aatgatacta acataacata gaacattttc agggaggacc ttgcttggag ggtaccgagc 9300  
tcagaaaaa 9309